

PS Disclosure; Page 73; 80pp; English.

XX The invention relates to nucleic acid molecules encoding non-aggregating
 CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drFP583 (NFP-6) mutant DNA of the invention
 XX SQ Sequence 678 BP; 145 A; 225 C; 203 G; 105 T; 0 U; 0 Other;

Query Match 100.0%; Score 678; DB 6; Length 678;
 Best Local Similarity 100.0%; Pred. No. 1.7e-109;
 Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60
 Db 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Qy 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTACGAGGGC 120
 Db 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTACGAGGGC 120

Qy 121 CACAACACCGTGAAAGTGAAGGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180
 Db 121 CACAACACCGTGAAAGTGAAGGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180

Qy 181 CTGTCCCGCCAGTTCAGTACGGCTCCAGGTTCAGGTGAAGCACCCCGCCGACATCCCC 240
 Db 181 CTGTCCCGCCAGTTCAGTACGGCTCCAGGTTCAGGTGAAGCACCCCGCCGACATCCCC 240

Qy 241 GACTACAGAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAATTCGAG 300
 Db 241 GACTACAGAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAATTCGAG 300

Qy 301 GACGGCGGTGGCGACCGGTGACCCAGGACTCTCTCCCTGCGAGGACGGCTGCTTCATCTAC 360
 Db 301 GACGGCGGTGGCGACCGGTGACCCAGGACTCTCTCCCTGCGAGGACGGCTGCTTCATCTAC 360

Qy 361 AAGTGAAGTTCATCGSCGTGAATTCCTCCGACGGCCCGCTGATGAGAGAGAGACC 420
 Db 361 AAGTGAAGTTCATCGSCGTGAATTCCTCCGACGGCCCGCTGATGAGAGAGAGACC 420

Qy 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGACCGCGCTGCTGAAGGGCGAG 480
 Db 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGACCGCGCTGCTGAAGGGCGAG 480

Qy 481 ATCCAAAGCCCTGAAGCTGAAGGACGGCGGCACCTACTGTGTGAGTTCAAGTCCATC 540
 Db 481 ATCCAAAGCCCTGAAGCTGAAGGACGGCGGCACCTACTGTGTGAGTTCAAGTCCATC 540

Qy 541 TACATGCCCAAGAACCGCGTGAAGTGGCGGCTACTACTGTGGACACCAAGCTGGAC 600
 Db 541 TACATGCCCAAGAACCGCGTGAAGTGGCGGCTACTACTGTGGACACCAAGCTGGAC 600

Qy 601 ATCACTTCCCAACAGGAGACTACCAATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660
 Db 601 ATCACTTCCCAACAGGAGACTACCAATCGTGGAGCAGTACGAGCGCACCGAGGGCCGC 660

Qy 661 CACCACCTGTCTCTGTAA 678
 Db 661 CACCACCTGTCTCTGTAA 678

RESULT 2
 AAD28209
 ID AAD28209 standard; cDNA; 678 BP.
 XX
 AC AAD28209;
 XX

DT 22-APR-2002 (first entry)

XX Discosoma sp. non-aggregating mutant timer protein ESNA encoding cDNA.

DE Fluorescent timer protein; protein movement; translocation; trafficking;
 KW promoter activity; gene expression; transgenic plant; gene modification;
 KW protein age; mutant; ESNA; ss.

XX Discosoma sp.
 OS Synthetic.

XX Key Location/Qualifiers
 FT 1..678 /*tag= a
 FT CDS /product= "Mutant ESNA protein"

XX WO200196373-A2.

XX 20-DEC-2001.

XX 13-JUN-2001; 2001WO-US019097.

XX 14-JUN-2000; 2000US-0211607P.

XX (CLON-) CLONTECH LAB INC.

XX Fradkov AF, Tersikh A;
 XX WPI: 2002-154595/20.
 XX P-PSDB; AAE17542.

XX New fluorescent timer proteins comprising an emission spectrum that
 changes over time from a first wavelength to a second wavelength, useful
 for monitoring intracellular protein movement, translocation, trafficking
 or stability.

XX Disclosure; Fig 3; 89pp; English.

XX The invention relates to a fluorescent timer protein having an emission
 spectrum that changes over time after synthesis from a first wavelength
 to a second wavelength. The fluorescent timer proteins are useful in
 CC monitoring the activity of a promoter, determining the age of a protein,
 CC identifying an agent that modulates the activity of a promoter and in
 CC enriching a population of cells comprising a fluorescent timer protein.
 CC The fluorescent timer proteins are also useful for assessing gene
 CC expression during development of a multicellular organism or during
 CC cellular differentiation, in response to a drug or other inducer of
 CC promoter activity, as a reporter to serve as a read-out of promoter
 CC activity, monitoring intracellular protein movement or translocation,
 CC protein trafficking, or protein stability, to investigate temporal
 CC aspects of the activity of a regulatory element, for determining cell
 CC fate during development and organ remodeling, in spatial and temporal
 CC visualisation of newly synthesised proteins and accumulated proteins, and
 CC in distinguishing between newly formed and pre-existing structures, e.g.
 CC membrane junctions and extracellular matrix components. The fluorescent
 CC timer proteins may further be used to investigate where photobleaching
 CC techniques are employed, as detectable labels, as selectable markers, as
 CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
 CC assays, and as second messenger detectors. The nucleic acids can be used
 CC to generate transgenic, non-human plants or animals or site-specific gene
 CC modifications in cell lines. The present sequence is a cDNA encoding
 CC Discosoma sp. ESNA non-aggregating fluorescent timer protein derived from
 CC fluorescent timer protein ES by substituting amino acids at positions
 CC R2A, K5E and K9T. ES protein is derived from humanised wild-type Anthozoa
 CC protein drFP583 by substituting Val to Ala at 105 and Ser to Thr at 197

XX SQ Sequence 678 BP; 145 A; 226 C; 203 G; 104 T; 0 U; 0 Other;

Query Match 99.8%; Score 676.4; DB 6; Length 678;
 Best Local Similarity 99.9%; Pred. No. 3.2e-109;
 Matches 677; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGC 60

Db 1 ATGGGCTCTCCGAGAACGTCATCACCGAGTTTCATGCGCTTCAAGGTGCGATGAGGGC 60
QY 61 ACCGTGAAGCGCCACGAGTTTCAGATCGAGGGCGAGGGCGGCCCTTACGAGGGC 120
Db 61 ACCGTGAAGCGCCACGAGTTTCAGATCGAGGGCGAGGGCGGCCCTTACGAGGGC 120
QY 121 CACAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCCCTTCCCTTGGGACATC 180
Db 121 CACAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCCCTTCCCTTGGGACATC 180
QY 181 CTGTCCCCCAGTTCCAGTACCGCTCCAGGTGTAGTGAAGCACCAGGACATCCCTCC 240
Db 181 CTGTCCCCCAGTTCCAGTACCGCTCCAGGTGTAGTGAAGCACCAGGACATCCCTCC 240
QY 241 GACTACAGAAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGGTGATGAATTCGAG 300
Db 241 GACTACAGAAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGGTGATGAATTCGAG 300
QY 301 GACGGGGCGTGGAGACCGTACCGAGACTCTCTCCCTGACGAGCGGTCTTCTATCTAC 360
Db 301 GACGGGGCGTGGAGACCGTACCGAGACTCTCTCCCTGACGAGCGGTCTTCTATCTAC 360
QY 361 AAGGTGAAGTTTCATGGCGTGAATTCCTCCCGAGGGCTTCAAGTGGAGCGGTGATGA 420
Db 361 AAGGTGAAGTTTCATGGCGTGAATTCCTCCCGAGGGCTTCAAGTGGAGCGGTGATGA 420
QY 421 ATGGGCTGGAGGCTTCCACGAGCGCTGTACCCCGCGAGCGGTGATGAAGGGCGAG 480
Db 421 ATGGGCTGGAGGCTTCCACGAGCGCTGTACCCCGCGAGCGGTGATGAAGGGCGAG 480
QY 481 ATCCACAAGCGCTGAAGCTGAAGCAGCGGGCCACTACTGTTGGAGTTCAAGTCCATC 540
Db 481 ATCCACAAGCGCTGAAGCTGAAGCAGCGGGCCACTACTGTTGGAGTTCAAGTCCATC 540
QY 541 TACATGGCCAAAGAGCGCGTGCAGCTGCCGGGTACTACTAGTGGACACCAAGCTGGAC 600
Db 541 TACATGGCCAAAGAGCGCGTGCAGCTGCCGGGTACTACTAGTGGACACCAAGCTGGAC 600
QY 601 ATCAGTCCCAACAGAGACTACACCATCGTGGAGCAGTACGAGGCGACCGAGGGCGGC 660
Db 601 ATCAGTCCCAACAGAGACTACACCATCGTGGAGCAGTACGAGGCGACCGAGGGCGGC 660
QY 661 CACCACTGTTCCTGTAA 678
Db 661 CACCACTGTTCCTGTAA 678

RESULT 3
AAD46281
ID AAD46281 standard; DNA; 675 BP.
XX AAD46281;
AC
XX
DT 27-DEC-2002 (first entry)
XX
DE Discosoma sp. drFP583 (NFP-6) mutant DNA, E57-NA.
XX
KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunsreen; second messenger detector; drFP583 protein; NFP-6; gene;
KW mutant; ds.
XX
OS Discosoma sp.
OS Synthetic.
XX
Key Location/Qualifiers
CDS 1..675
FT /*tag= a
FT /product= "drFP583 mutant protein, E57-NA"
FT /note= "No stop codon"

/partial
XX WO200268459-A2.
PN
XX
PD 06-SEP-2002.
XX
PF 20-FEB-2002; 2002WO-US005749.
XX
XX 21-FEB-2001; 2001US-0270983P.
PR 04-DEC-2001; 2001US-00006922.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
FI
XX
XX WPI; 2002-691654/74.
DR
XX
XX P-PSDB; AAE28836.
XX
PT New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
PT of an aggregating Chidarian chromo- or fluorescent protein or mutant for
PT analyte detection assays or fluorescence activated cell sorting
PT applications.
XX
XX Disclosure; Page 73; 80pp; English.
XX
XX The invention relates to nucleic acid molecules encoding non-aggregating
XX chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
XX useful in analyte detection assays, as colouring agents, as markers in
XX recombinant DNA applications, as sunsreens or filters, in fluorescence
XX resonance energy transfer (FRET) applications, as biosensors in
XX prokaryotic and eukaryotic cells, in screening assays, as second
XX messenger detectors, in fluorescence activated cell sorting applications,
XX in protease cleavage assays or as fluorescent timers. The present
XX sequence is Discosoma sp. drFP583 (NFP-6) mutant DNA of the invention
XX
SQ Sequence 675 BP; 142 A; 227 C; 204 G; 102 T; 0 U; 0 Other;

Query Match 98.8%; Score 670.2; DB 6; Length 675;
Best Local Similarity 99.6%; Pred. No. 3.8e-108;
Matches 672; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGCTCTCCGAGAGAGTCATCACCGAGTTTCATGCGCTTCAAGGTGCGCATGGAGGC 60
Db 1 ATGGCTCTCCGAGAGAGTCATCACCGAGTTTCATGCGCTTCAAGGTGCGCATGGAGGC 60
QY 61 ACCGTGAAGCGCCACGAGTTTCAGATCGAGGGCGAGGGCGGCCCTTACGAGGGC 120
Db 61 ACCGTGAAGCGCCACGAGTTTCAGATCGAGGGCGAGGGCGGCCCTTACGAGGGC 120
QY 121 CACAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCCCTTCCCTTGGGACATC 180
Db 121 CACAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCCCTTCCCTTGGGACATC 180
QY 181 CTGTCCCCCAGTTTCAGTACCGCTCCAGGTGTACGTGAAGCACCAGGCGCATCCCC 240
Db 181 CTGTCCCCCAGTTTCAGTACCGCTCCAGGTGTACGTGAAGCACCAGGCGCATCCCC 240
QY 241 GACTACAGAAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATTCGAG 300
Db 241 GACTACAGAAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATTCGAG 300
QY 301 GACGGGGCGTGGAGACCGTACCGAGACTCTCTCCCTGACGAGCGGTCTTCTATCTAC 360
Db 301 GACGGGGCGTGGAGACCGTACCGAGACTCTCTCCCTGACGAGCGGTCTTCTATCTAC 360
QY 361 AAGGTGAAGTTTCATGGCGTGAATTCCTCCCGAGGGCTTCAAGTGGAGCGCGTGAAG 420
Db 361 AAGGTGAAGTTTCATGGCGTGAATTCCTCCCGAGGGCTTCAAGTGGAGCGCGTGAAG 420
QY 421 ATGGGCTGGAGGCTTCCACGAGCGCTGTACCCCGCGAGCGGTGATGAAGGGCGAG 480
Db 421 ATGGGCTGGAGGCTTCCACGAGCGCTGTACCCCGCGAGCGGTGATGAAGGGCGAG 480

QY 481 ATCCACAGGCCCTGAAGCTGAAGAGCGGGCCACCTACCTGGTGGAGTTCAAGTCCATC 540
 Dd |||||
 QY 481 ACCCACAGGCCCTGAAGCTGAAGAGCGGGCCACCTACCTGGTGGAGTTCAAGTCCATC 540
 Dd |||||
 QY 541 TACATGGCCAAAGAGCCCGTGCAGTCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
 Dd |||||
 QY 541 TACATGGCCAAAGAGCCCGTGCAGTCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
 Dd |||||
 QY 601 ATCACTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
 Dd |||||
 QY 601 ATCACTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
 Dd |||||
 QY 661 CACCACCTGTTCCTG 675
 Dd |||||
 QY 661 CACCACCTGTTCCTG 675

RESULT 4

AAD28208
 ID AAD28208 standard; cDNA; 678 BP.

XX AAD28208;
 AC

XX 22-APR-2002 (first entry)
 DT

XX Discosoma sp. fluorescent protein E5 encoding cDNA.
 DE

XX Fluorescent timer protein; protein movement; translocation; trafficking;
 KW promoter activity; gene expression; transgenic protein; gene modification;
 KW protein age; E5; ss.
 KW

XX Discosoma sp.
 OS

XX Key Location/Qualifiers
 FH 1. 678
 CDS /*tag= a
 FT /product= "E5 protein"
 FT

XX WO200196373-A2.
 PN

XX 20-DEC-2001.
 PD

XX 13-JUN-2001; 2001WO-US019097.
 XX

XX 14-JUN-2000; 2000US-0211607P.
 XX

XX (CLON-) CLONTECH LAB INC.
 XX

XX Fradkov AF, Tersikh A;
 XX

XX WPI: 2002-154595/20.
 XX

XX P-PSDB; AAE17541.
 XX

XX New fluorescent timer proteins comprising an emission spectrum that
 PT changes over time from a first wavelength to a second wavelength, useful
 PT for monitoring intracellular protein movement, translocation, trafficking
 PT or stability.
 PT

XX Disclosure; Fig 2; 89pp; English.
 XX

XX The invention relates to a fluorescent timer protein having an emission
 CC spectrum that changes over time after synthesis from a first wavelength
 CC to a second wavelength. The fluorescent timer proteins are useful in
 CC monitoring the activity of a promoter, determining the age of a protein,
 CC identifying an agent that modulates the activity of a promoter and in
 CC enriching a population of cells comprising a fluorescent timer protein.
 CC The fluorescent timer proteins are also useful for assessing gene
 CC expression during development of a multicellular organism or during
 CC cellular differentiation, in response to a drug or other inducer of
 CC promoter activity, as a reporter to serve as a read-out of promoter
 CC activity, monitoring intracellular protein movement or translocation,
 CC protein trafficking, or protein stability, to investigate temporal
 CC aspects of the activity of a regulatory element, for determining cell

CC fate during development and organ remodelling, in spatial and temporal
 CC visualisation of newly synthesised proteins and accumulated proteins, and
 CC in distinguishing between newly formed and pre-existing structures, e.g.
 CC membrane junctions and extracellular matrix components. The fluorescent
 CC timer proteins may further be used to investigate where photobleaching
 CC techniques are employed, as detectable labels, as selectable markers, as
 CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
 CC assays, and as second messenger detectors. The nucleic acids can be used
 CC to generate transgenic, non-human plants or animals or site-specific gene
 CC modifications in cell lines. The present sequence is a cDNA encoding
 CC Discosoma sp. E5 fluorescent protein derived from humanised wild-type
 CC Anthonia protein drFP593 by substituting Val to Ala at 105 and Ser to Thr
 CC at 197
 XX
 SQ Sequence 678 BP; 147 A; 224 C; 203 G; 104 T; 0 U; 0 Other;

Query Match 98.6%; Score 668.4; DB 6; Length 678;

Best Local Similarity 99.1%; Pred. No. 7.8e-108; Indels 0; Gaps 0;

Matches 672; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGC 60
 Dd |||||
 QY 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTCATGCGCTTCAAGGTGCGCATGGAGGC 60
 Dd |||||
 QY 61 ACCGTGAAGCCACGAGATTGAGATCGAGGCGAGGCGGCGCCCTCTAGAGGC 120
 Dd |||||
 QY 61 ACCGTGAAGCCACGAGATTGAGATCGAGGCGAGGCGGCGCCCTCTAGAGGC 120
 Dd |||||
 QY 121 CACAACACCGTGAAGTTCAGAGTGACCAAGGCGGCGCCCTCTCGCTTGGGACATC 180
 Dd |||||
 QY 121 CACAACACCGTGAAGTTCAGAGTGACCAAGGCGGCGCCCTCTCGCTTGGGACATC 180
 Dd |||||
 QY 181 CTGTCCCTCCAGTTCAGATCGGCTCCAGGTGATGAGACACCCCGCGGACATCCCC 240
 Dd |||||
 QY 181 CTGTCCCTCCAGTTCAGATCGGCTCCAGGTGATGAGACACCCCGCGGACATCCCC 240
 Dd |||||
 QY 241 GACTACAGAGAGCTGCTCTCCCGAGGCTTCAAGTGGGAGCGCGTGATGAACCTTCGAG 300
 Dd |||||
 QY 241 GACTACAGAGAGCTGCTCTCCCGAGGCTTCAAGTGGGAGCGCGTGATGAACCTTCGAG 300
 Dd |||||
 QY 301 GACGCGCGGTGGCGACCGTGACCCAGGACTCTCTCCCTGAGGACGCGCTCTTCAATC 360
 Dd |||||
 QY 301 GACGCGCGGTGGCGACCGTGACCCAGGACTCTCTCCCTGAGGACGCGCTCTTCAATC 360
 Dd |||||
 QY 361 AAGGTGAAGTTCATCGGCGTGAACCTCTCCCTCGAGCGGCGCGTGATGCGAGAGAGACC 420
 Dd |||||
 QY 361 AAGGTGAAGTTCATCGGCGTGAACCTCTCCCTCGAGCGGCGCGTGATGCGAGAGAGACC 420
 Dd |||||
 QY 421 ATGGGCTGGGAGGCTCTCACCGAGCGCTGTACCCCGCGGACGCGCTGTGAAGGGCGAG 480
 Dd |||||
 QY 421 ATGGGCTGGGAGGCTCTCACCGAGCGCTGTACCCCGCGGACGCGCTGTGAAGGGCGAG 480
 Dd |||||
 QY 481 ATCCACAGAGCCCTGAAGCTGAAGAGCGGCGGCGGCGGCGGCTGAGTTCAGTCCATC 540
 Dd |||||
 QY 481 ATCCACAGAGCCCTGAAGCTGAAGAGCGGCGGCGGCGGCGGCTGAGTTCAGTCCATC 540
 Dd |||||
 QY 541 TACATGGCCAAAGAGCCCGTGCAGTCCCGGCTACTACTACGTGGGACACCAAGCTGGAC 600
 Dd |||||
 QY 541 TACATGGCCAAAGAGCCCGTGCAGTCCCGGCTACTACTACGTGGGACACCAAGCTGGAC 600
 Dd |||||
 QY 601 ATCACTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
 Dd |||||
 QY 601 ATCACTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
 Dd |||||
 QY 661 CACCACCTGTTCCTGTA 678
 Dd |||||
 QY 661 CACCACCTGTTCCTGTA 678

RESULT 5

AAD46278
 ID AAD46278 standard; DNA; 678 BP.

XX

AC	AAD46278;	Db	181	CTGTCCCTCCAGTTCACAGTACGCTCCAAAGTGTACGTGAAGCACCCCGCCGACATCCCC	240
XX					
DT	27-DEC-2002 (first entry)	Qy	241	GACTACAAGAAGCTGTCTTCCCTCCCGAGGGGCTTCAAGTGGAGCGCGTGATGAATCTCGAG	300
XX					
DE	Discosoma sp. drFP583 (NFP-6) wild-type protein encoding DNA.	Db	241	GACTACAAGAAGCTGTCTTCCCTCCCGAGGGGCTTCAAGTGGAGCGCGTGATGAATCTCGAG	300
XX					
KW	Fluorescent protein; chromoprotein; protease cleavage assay; filter;	Qy	301	GAGGGGGCTGCGAGCCGCTGACCCAGGACTCTCCCTCCAGGACCGCTCTTCACTCTAC	360
KW	fluorescence activated cell sorting application; fluorescent timer;				
KW	biosensor; fluorescence resonance energy transfer application; FRET;	Db	301	GAGGGGGCTGCTGACCCGCTGACCCAGGACTCTCTCCCTCCAGGACCGCTCTTCACTCTAC	360
KW	colouring agent; recombinant DNA application; analyte detection assay;	Qy	361	AAGGTGAAGTTTCATCGGCGTGAATCTTCCCTCCAGGCGGCCCTGATGCAAGAAGACC	420
KW	screen; second messenger detector; drFP583 protein; NFP-6; gene; ds.				
XX		Db	361	AAGGTGAAGTTTCATCGGCGTGAATCTTCCCTCCAGGCGGCCCTGATGCAAGAAGACC	420
OS	Discosoma sp.				
XX		Qy	421	ATGGGCTGGAGGCTTCCACCGAGCGCTTACCCCGCGACCGCGCTGTGAAGGGCGAG	480
FH	Key				
FT	CDS	Db	421	ATGGGCTGGAGGCTTCCACCGAGCGCTTACCCCGCGACCGCGCTGTGAAGGGCGAG	480
FT					
FT	/product= "drFP583 wild-type protein"				
XX		Qy	481	ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCACCTACCTGTGGAGTTCAAGTCCATC	540
XX	WO200269459-A2.	Db	481	ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCACCTACCTGTGGAGTTCAAGTCCATC	540
XX		Qy	541	TACATGGCCCAAGAGCCGCTGCAGCTGCCGGCTACTACTACGTGGACACCAAGCTGGAC	600
XX	06-SEP-2002.	Db	541	TACATGGCCCAAGAGCCGCTGCAGCTGCCGGCTACTACTACGTGGACACCAAGCTGGAC	600
XX	20-FEB-2002; 2002WO-US005749.	Qy	601	ATCACCTCCACACGAGGACTACACCATCTGTGAGCAGTACGAGCGCACCGAGGGCGGC	660
XX	21-FEB-2001; 2001US-0270983P.	Db	601	ATCACCTCCACACGAGGACTACACCATCTGTGAGCAGTACGAGCGCACCGAGGGCGGC	660
XX	04-DEC-2001; 2001US-00006922.	Qy	661	CACCACCTGTTCTCTGTAA 678	
XX	(CLON-) CLONTECH LAB INC.	Db	661	CACCACCTGTTCTCTGTAA 678	
XX	Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;				
XX	WPI; 2002-691654/74.				
DR	P-PSDB; AAE28833.				
XX					
PT	New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant				
PT	of an aggregating Cnidarian chromo- or fluorescent protein or mutant for				
PT	analyte detection assays or fluorescence activated cell sorting				
PT	applications.				
XX					
PS	Disclosure; Page 70; 80pp; English.				
XX					
CC	The invention relates to nucleic acid molecules encoding non-aggregating				
CC	chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are				
CC	useful in analyte detection assays, as colouring agents, as markers in				
CC	recombinant DNA applications, as sunscreens or filters, in fluorescence				
CC	resonance energy transfer (FRET) applications, as biosensors in				
CC	prokaryotic and eukaryotic cells, in screening assays, as second				
CC	messenger detectors, in fluorescence activated cell sorting applications,				
CC	in protease cleavage assays or as fluorescent timers. The present				
CC	sequence is a DNA encoding Discosoma sp. drFP583 (NFP-6) wild-type				
CC	protein of the invention				
XX					
SQ	Sequence 678 BP; 146 A; 223 C; 203 G; 106 T; 0 U; 0 Other;				
	Query Match 98.1%; Score 665.2; DB 6; Length 678;				
	Best Local Similarity 98.8%; Pred. No. 2.8e-107;				
	Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;				
Qy	1 ATGGCTCTCCGAGAACGTCATCACCGAGTTTCATGGCTTCAAGTGGCATGAGGGC 60				
Db	1 ATGGCTCTCCGAGAACGTCATCACCGAGTTTCATGGCTTCAAGTGGCATGAGGGC 60				
Qy	61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGGCGAGGGCGGCCCTACGAGGGC 120				
Db	61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGGCGAGGGCGGCCCTACGAGGGC 120				
Qy	121 CACAACACCGTGAAGTTGAAGGTGACCAAGGGCGGCCCTGCGCTGGGACATC 180				
Db	121 CACAACACCGTGAAGTTGAAGGTGACCAAGGGCGGCCCTGCGCTGGGACATC 180				
Qy	181 CTGTCCCTCCAGTTCACAGTACGCTCCAAAGTGTACGTGAAGCACCCCGCCGACATCCCC				

New fluorescent timer proteins comprising an emission spectrum that changes over time from a first wavelength to a second wavelength, useful

PT for monitoring intracellular protein movement, translocation, trafficking
 XX or stability.

PS Example 1; Fig 1; 89pp; English.

XX The invention relates to a fluorescent timer protein having an emission
 CC spectrum that changes over time after synthesis from a first wavelength
 CC to a second wavelength. The fluorescent timer proteins are useful in
 CC monitoring the activity of a promoter, determining the age of a protein,
 CC identifying an agent that modulates the activity of a promoter and in
 CC enriching a population of cells comprising a fluorescent timer protein.
 CC The fluorescent timer proteins are also useful for assessing gene
 CC expression during development of a multicellular organism or during
 CC cellular differentiation, in response to a drug or other inducer of
 CC promoter activity, as a reporter to serve as a read-out of promoter
 CC activity, monitoring intracellular protein movement or translocation,
 CC protein trafficking, or protein stability, to investigate temporal
 CC aspects of the activity of a regulatory element, for determining cell
 CC fate during development and organ remodeling, in spatial and temporal
 CC visualisation of newly synthesised proteins and accumulated proteins, and
 CC in distinguishing between newly formed and pre-existing structures, e.g.
 CC membrane junctions and extracellular matrix components. The fluorescent
 CC timer proteins may further be used to investigate where photobleaching
 CC techniques are employed, as detectable labels, as selectable markers, as
 CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
 CC assays, and as second messenger detectors. The nucleic acids can be used
 CC to generate transgenic, non-human plants or animals or site-specific gene
 CC modifications in cell lines. The present sequence is a DNA encoding
 CC Discosoma sp. humanised wild-type Anthozoa protein drFP583 used for
 CC generating fluorescent proteins

XX Sequence 678 BP; 146 A; 233 C; 203 G; 106 T; 0 U; 0 Other;

Query Match 98.1%; Score 665.2; DB 6; Length 678;

Best Local Similarity 98.8%; Pred. No. 2.8e-107;

Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGGCTCTCCGAGACGTCATCCAGGTTTCATGGCTTCAAGTGGCGCATGGAGGC 60
 DB 1 ATGGCTCTCCGAGACGTCATCCAGGTTTCATGGCTTCAAGTGGCGCATGGAGGC 60
 QY 61 ACCGTGAACGCCACAGGTTTCGAGATCGAGGCGAGGCGGCGGCGGCGGCGGCGG 120
 DB 61 ACCGTGAACGCCACAGGTTTCGAGATCGAGGCGAGGCGGCGGCGGCGGCGGCGG 120
 QY 121 CACACACCGTGAAGTTGAAGTTGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 DB 121 CACACACCGTGAAGTTGAAGTTGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 QY 181 CTGTCGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 DB 181 CTGTCGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 QY 241 GACTACAAAGAGTGTCTTCCCGAGGCGTTCAGTGGGAGCGGTGATGAATTCGAG 300
 DB 241 GACTACAAAGAGTGTCTTCCCGAGGCGTTCAGTGGGAGCGGTGATGAATTCGAG 300
 QY 301 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
 DB 301 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
 QY 361 AAGTGAAGTTTCATCGGCGTGAATTCCTCCCGAGGCGGCGGCGGCGGCGGCGG 420
 DB 361 AAGTGAAGTTTCATCGGCGTGAATTCCTCCCGAGGCGGCGGCGGCGGCGGCGG 420
 QY 421 ATGGGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
 DB 421 ATGGGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
 QY 481 ATCCACAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
 DB 481 ATCCACAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540

QY 541 TACATGGCCAAAGACCCGTCAGCTGCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
 DB 541 TACATGGCCAAAGACCCGTCAGCTGCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
 QY 601 ATCACTCTCCCAACAGGAGTACACCATCTGTGGAGCGAGTACGAGCGGACCCGAGGGCGC 660
 DB 601 ATCACTCTCCCAACAGGAGTACACCATCTGTGGAGCGAGTACGAGCGGACCCGAGGGCGC 660
 QY 661 CACCACTGTTCCTGTAA 678
 DB 661 CACCACTGTTCCTGTAA 678

RESULT 7

AAA48743

ID AAA48743 standard; cDNA; 695 BP.

XX AAA48743;

AC AAA48743;

XX 19-SEP-2000 (first entry)

XX Humanised Discosoma sp. "red" novel fluorescent protein drFP583 cDNA.

XX Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;

XX fluorescent labeling; ss.

XX Discosoma sp; "red".

OS Synthetic.

XX WO200034326-A1.

XX 15-JUN-2000.

XX 10-DEC-1999; 99WO-US029473.

XX 11-DEC-1998; 98US-00210330.

PR 14-OCT-1999; 99US-00418529.

XX (CLON-) CLONTECH LAB INC.

PI Lukyanov SA, Fradkov AP, Labas YA, Matz MV, Green G, Chen Y;

PI Ding Li;

XX WPI; 2000-423381/36.

XX Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
 PT useful for fluorescent labeling and as markers.

XX Claim 6; Page 75-76; 86pp; English.

XX The present sequence is humanised drFP583 cDNA. drFP583 is a full-length
 CC cDNA encoding a novel fluorescent protein (nFP) from Discosoma sp. "red",
 CC a non-bioluminescent species of the Class Anthozoa. The wild-type drFP583
 CC nucleotide sequence was altered to optimise the codons for expression of
 CC the fluorescent protein in mammalian cells. Fluorescent proteins can be
 CC used in fluorescent labeling, a useful tool for marking a protein, cell
 CC or organism of interest. Unlike other markers used in protein labeling,
 CC such as beta-galactosidase and luciferase, fluorescent proteins do not
 CC require an exogenous cofactor or substrate. Methods involving fluorescent
 CC proteins are also less laborious and less difficult to control than the
 CC traditional methods of fluorescent labeling, where a protein of interest
 CC is purified and then covalently conjugated to a fluorophore derivative.
 CC Novel fluorescent proteins isolated from species of the Class Anthozoa
 CC can be used as markers for gene expression and protein localization
 CC studies, and in fluorescence resonance energy transfer (FRET) reactions.
 CC They may have improved properties and better suitability for larger
 CC fluorescent protein

XX Sequence 695 BP; 149 A; 228 C; 209 G; 109 T; 0 U; 0 Other;

Query Match 98.1%; Score 665.2; DB 3; Length 695;

Best Local Similarity 98.8%; Pred. No. 2.8e-107;

Db 245 ACTACAAGAGCTGTCTTCCCTCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG 304
 QY 302 ACGGGCGGTGGACCGTGAACCCAGGACTCTCCCTGCGAGGACGGTCTTCACTACA 361
 Db 305 ACGGCGGCGTGGTACCGTGAACCCAGGACTCTCCCTGCGAGGACGGTCTTCACTACA 364
 QY 362 AGGTCAAGTTCATCGGGTGAACCTTCCCTCCGAGGGCTTCAAGTGGGAGCGGTGATGACGAAGAAGACCA 421
 Db 365 AGGTGAAGTTCATCGGGTGAACCTTCCCTCCGAGGGCTTCAAGTGGGAGCGGTGATGACGAAGAAGACCA 424
 QY 422 TGGGCTGGGAGGCTTCAACGAGGCGCTGTACCCCGCGGACGGGCTGCTGAAGGGCGAGA 481
 Db 425 TGGGCTGGGAGGCTTCAACGAGGCGCTGTACCCCGCGGACGGGCTGCTGAAGGGCGAGA 484
 QY 482 TCCACAAGGCGCTGAAGCTGAAGGAGCGGCGCACTACCTGGTGGAGTTCAAGTCCATCT 541
 Db 485 TCCACAAGGCGCTGAAGCTGAAGGAGCGGCGCACTACCTGGTGGAGTTCAAGTCCATCT 544
 QY 542 ACATGGCAAGAGCCCGTGAAGTGAAGGAGCGGCGCACTACCTGGTGGAGTTCAAGTCCATCT 601
 Db 545 ACATGGCAAGAGCCCGTGAAGTGAAGGAGCGGCGCACTACCTGGTGGAGTTCAAGTCCATCT 604
 QY 602 TCACCTCCCAACGAGGAGTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGC 661
 Db 605 TCACCTCCCAACGAGGAGTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGC 664
 QY 662 ACCACCTGTTCCCTGTA 677
 Db 665 ACCACCTGTTCCCTGTA 680

RESULT 9
 AAL47954
 ID AAL47954 standard; DNA; 4692 BP.
 XX AC AAL47954;
 DT 26-SEP-2002 (first entry)
 XX DE Modified yeast strain related vector pDsRed1-N1.
 XX KW Yeast; RAD54; promoter; genotoxicity cassette; cytotoxicity cassette;
 XX KW modified yeast strain; environmental pollution; vector; ds.
 XX OS Synthetic.
 XX FN DE10061872-A1.
 XX PD 20-JUN-2002.
 XX PF 12-DEC-2000; 2000DE-01061872.
 XX PR 12-DEC-2000; 2000DE-01061872.
 XX PA (LICH/) LICHTENBERG-FRATE H.
 XX PI Lichtenberg-Frate H;
 XX DR WPI; 2002-539633/58.
 XX PT Modified yeast strain, useful for detecting toxic compounds in
 XX PT environment, contains integrated cassettes responsive to genotoxic and
 XX PT cytotoxic compounds.
 XX PS Example 1; Page 24-25; 34pp; German.
 XX CC The present invention relates to a modified yeast strain that contains,
 CC integrated stably and functionally in its genome, a genotoxicity cassette
 CC and a cytotoxicity cassette, each comprising a promoter and reporter
 CC gene, both of which are different in the two cassettes. The modified
 CC yeast strain is used to detect environmental pollution, especially
 CC genotoxic and/or cytotoxic substances in complex environmental
 CC contaminants, especially organic compounds, but also (non-)ionising

CC radiation and chemical carcinogens. Particular applications are in
 CC monitoring (waste) water (e.g. as an early warning system), medical
 CC toxicology screening and for industrial process control. The present
 CC sequence is a vector suitable for use in the present invention
 XX
 SQ Sequence 4692 BP; 1111 A; 1268 C; 1247 G; 1066 T; 0 U; 0 Other;
 Query Match 97.6%; Score 661.6; DB 6; Length 4692;
 Best Local Similarity 98.7%; Pred. No. 1.1e-106;
 Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 2 TGGGCTCTCCGAGAAAGTTCATACCGAGTTTCATCGGCTTCAAGTGGGAGCGGTGATGACGAAGAAGACCA 61
 Db 683 TGGGCTCTCCGAGAAAGTTCATACCGAGTTTCATCGGCTTCAAGTGGGAGCGGTGATGACGAAGAAGACCA 742
 QY 62 CCGTGAACGGCCACGAGTTTCGAGATCGAGGCGGAGGGCGGCGGCTTACGAGGCGCC 121
 Db 743 CCGTGAACGGCCACGAGTTTCGAGATCGAGGCGGAGGGCGGCGGCTTACGAGGCGCC 802
 QY 122 ACAACACCGTGAAGTTGAAGGTGACCAAGGCGGCGGCGGCTTCCCTTCCCTGGGACATCC 181
 Db 803 ACAACACCGTGAAGTTGAAGGTGACCAAGGCGGCGGCGGCTTCCCTTCCCTGGGACATCC 862
 QY 182 TGTCCCCCAGTTTCAGTACCGCTTCCAGGTGACGTGAGGAGCGGCGGCGGCTTCCCTGGGACATCC 241
 Db 863 TGTCCCCCAGTTTCAGTACCGCTTCCAGGTGACGTGAGGAGCGGCGGCGGCTTCCCTGGGACATCC 922
 QY 242 ACTACAAGAAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG 301
 Db 923 ACTACAAGAAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG 982
 QY 302 ACGGCGGCTGGGACCGTGAACCGAGTCTCTCCCTGCGAGGACGGCTGCTTCACTACA 361
 Db 983 ACGGCGGCTGGTGAACCGTGAACCGAGTCTCTCCCTGCGAGGACGGCTGCTTCACTACA 1042
 QY 362 AGGTGAAGTTCATCGGGTGAACCTTCCCTCCGAGGCGGCGGTGATGACGAAGAAGACCA 421
 Db 1043 AGGTGAAGTTCATCGGGTGAACCTTCCCTCCGAGGCGGCGGTGATGACGAAGAAGACCA 1102
 QY 422 TGGGCTGGGAGGCTTCCACCGAGCGCTGTACCCCGCGAGCGGCTGCTGAAAGGGGAGA 481
 Db 1103 TGGGCTGGGAGGCTTCCACCGAGCGCTGTACCCCGCGAGCGGCTGCTGAAAGGGGAGA 1162
 QY 482 TCCACAGGCGCTGAAGTGAAGGAGCGGCGGCGCACTACCTGCTGGAGTTCAGTCCATCT 541
 Db 1163 TCCACAGGCGCTGAAGTGAAGGAGCGGCGGCGCACTACCTGCTGGAGTTCAGTCCATCT 1222
 QY 542 ACATGGGCAAGAAGCGGCTGAGCTGCGGCGCTACTACTACGTGGACACCAAGCTGGACA 601
 Db 1223 ACATGGGCAAGAAGCGGCTGAGCTGCGGCGCTACTACTACGTGGACTTCAAGCTGGACA 1282
 QY 602 TCACCTGCCACACGAGGAGTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 661
 Db 1283 TCACCTGCCACACGAGGAGTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 1342
 QY 662 ACCACCTGTTCCCTGTA 677
 Db 1343 ACCACCTGTTCCCTGTA 1358

RESULT 10
 ACC44640
 ID ACC44640 standard; DNA; 4692 BP.
 XX AC ACC44640;
 XX DT 29-MAY-2003 (first entry)
 XX DE Vector pDsRed1N1 nucleotide sequence SEQ ID NO:29.
 XX KW Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
 KW att site; integrase; recombinase; ACes; gene therapy; transgenic animal;
 KW platform artificial chromosome expression system; gene; ds.

CC control over fertility and reproduction. The construct comprises a native
CC promoter, a blocking DNA sequence contoured for and designed to abrogate
CC a crucial gene's function or to cause its mis-expression, and a genetic
CC switch to regulate controlled expression/repression of the blocker/gene
CC knockout. The construct is useful for preventing embryogenesis or
CC gametogenesis in animals by stably transforming an animal cell with the
CC construct by microinjection, transfection or infection, where the
CC construct stably integrates into the genome by homologous recombination,
CC and implanting the cell into a host organism, where a whole animal
CC develops from the implanted cell. The present DNA sequence is plasmid
CC construct pBir(DHSP)-RFP-ORFDS/SH used for transfection of Pacific
CC oysters. The plasmid comprises oyster Hox gene under the under the
CC control of Drosophila heat shock promoter (dHSP), red fluorescent protein
CC and SV40 PolyA
XX
SQ Sequence 7910 BP; 2016 A; 1893 C; 1924 G; 2077 T; 0 U; 0 Other;
Query Match 97.68; Score 661.6; DB 4; Length 7910;
Best Local Similarity 98.74; Pred. No. 1.1e-106;
Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
2 TGGCTCTCTCGAGAACGTCATCACCGAGTTTCATGGCGTTTCAAGTGCGCATGGAGGCA 61
2704 TGGCTCTCTCGAGAACGTCATCACCGAGTTTCATGGCGTTTCAAGTGCGCATGGAGGCA 2645
62 CCGTGAACGCGCACGAGTTTCAGATCGAGGCGAGGCGGAGGCGCCCGCTACGAGGGCC 121
2644 CCGTGAACGCGCACGAGTTTCAGATCGAGGCGAGGCGGAGGCGCCCGCTACGAGGGCC 2585
122 ACACACCGTGAAGTTTCAGATCGAGGCGGAGGCGGAGGCGCCCGCTACGAGGGCC 181
2584 ACACACCGTGAAGTTTCAGATCGAGGCGGAGGCGGAGGCGCCCGCTACGAGGGCC 2525
182 TGTCCTCCCGAGTTTCAGATCGAGGCGGAGGCGGAGGCGCCCGCTACGAGGGCC 241
2524 TGTCCTCCCGAGTTTCAGATCGAGGCGGAGGCGGAGGCGCCCGCTACGAGGGCC 2465
242 ACTACAGAGAGCTGCTCTCCCGAGGCTTCAGTGGGAGCGGTGATGAGTTCGAGG 301
2464 ACTACAGAGAGCTGCTCTCCCGAGGCTTCAGTGGGAGCGGTGATGAGTTCGAGG 2405
302 ACGGCGCGGTGGCGACCGTGCACCGAGGCTCTCCCGTGCAGGAGCGGCTGCTTCACTACA 361
2404 ACGGCGCGGTGGCGACCGTGCACCGAGGCTCTCCCGTGCAGGAGCGGCTGCTTCACTACA 2345
362 AGGTGAAGTTTCATCGGCGTGAAGTTTCCTCCGACGCGCCCGGTGATGAGGAGGACCA 421
2344 AGGTGAAGTTTCATCGGCGTGAAGTTTCCTCCGACGCGCCCGGTGATGAGGAGGACCA 2285
422 TGGGCTGGGAGGCTCTCCCGAGGCTTCAGTGGGAGCGGTGATGAGTTCGAGG 481
2284 TGGGCTGGGAGGCTCTCCCGAGGCTTCAGTGGGAGCGGTGATGAGTTCGAGG 2225
482 TCACAAAGGCGCTGAAGTTTCAGGAGCGGCGGACCTACCTGGTGGAGTTCAAGTCCATCT 541
2224 TCACAAAGGCGCTGAAGTTTCAGGAGCGGCGGACCTACCTGGTGGAGTTCAAGTCCATCT 2165
542 ACATGGCCAGAGCGGCTGAGTGGCGGCTGCTGAGGAGCGGCTGAGGAGCGGAGCA 601
2164 ACATGGCCAGAGCGGCTGAGTGGCGGCTGCTGAGGAGCGGCTGAGGAGCGGAGCA 2105
602 TCACCTCCCAACAGGAGGACTTACACCTGCTGGAGGAGTACGAGCGGCGGCGGCGCC 661
2104 TCACCTCCCAACAGGAGGACTTACACCTGCTGGAGGAGTACGAGCGGCGGCGGCGCC 2045
662 ACCACCTGTTCTGTA 677
2044 ACCACCTGTTCTGTA 2029
RESULT 12
ABS56664
ID ABS56664 standard; DNA; 9320 BP.

XX
AC ABS56664;
XX
DT 24-JAN-2003 (first entry)
XX
DE Plasmid p5-Puro-CMV-(N'-EGFP)-CMV-Red(EGFP-EJ) DNA.
XX
KW Vector; genotoxicity; food component; cosmetic; medicine; cancer;
XX genetic stability; hygromycin; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1..1592
FT /tag= a
FT /note= "retroviral vector p5NM"
FT CDS 1617..2216
FT /tag= b
FT /product= "puromycin resistance gene from pRetroOn"
FT promoter 2267..2848
FT /tag= c
FT /note= "CMV promoter from pEGFP-N1"
FT CDS 2906..3348
FT /tag= d
FT /product= "N'-EGFP"
FT misc_feature 3374..3392
FT /tag= e
FT /note= "retroviral vector p5NM"
FT promoter 3411..3992
FT /tag= f
FT /note= "CMV promoter from pDsRed1-N1"
FT CDS 4038..4718
FT /tag= g
FT /product= "pDsRed1-N1"
FT CDS 4766..5508
FT /tag= h
FT /product= "EGFP-EJ"
FT misc_feature 5527..9320
FT /tag= i
FT /note= "retroviral vector p5NM"

WO200270740-A2.

12-SEP-2002.

28-FEB-2002; 2002WO-EP002194.

05-MAR-2001; 2001DE-01010449.

(WIES/) WIESMUELLER L.

Wiesmuller L;

WPI; 2002-682909/73.

New vector containing recombination sequences, useful e.g. for assessing compounds for genotoxicity, cancer susceptibility and mutational frequency.

Disclosure; Page 104-106; 106pp; German.

This invention describes a novel vector containing, at least, two homologous DNA sequences that, through DNA exchange, induce a change in a gene present in the vector. The products of the invention are used to assess genotoxicity of compounds (e.g. components of foods, cosmetics and medicines), to determine susceptibility for development or progression of cancer, for analysis of (non-)conservative recombination processes, and joining and inactivating/reverting mutations, for characterization of selected genes with respect to maintenance or reduction of genetic stability, and to determine genetic (in)stability and recombinational or mutational frequency of cells, tissues or organisms. The vector allows an analysis to be performed in hours (making it suitable for large-scale, routine use), and can be used with a variety of eukaryotic cells and

CC animals. All known types of DNA recombinations can be examined. This
 CC sequence represents a plasmid p5-Puro-CMV- (N'-EGFP)-CMV-Red(EGFP-EU) DNA,
 CC described in the disclosure of the invention
 XX
 SQ Sequence 9320 BP; 2124 A; 2687 C; 2474 G; 2035 T; 0 U; 0 Other;

Query Match 97.6%; Score 661.6; DB 6; Length 9320;
 Best Local Similarity 98.7%; Pred. No. 1.1e-106;
 Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGCTCTCCCGAGACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGCA 61
 Db TGGCTCTCTCCAGAACGTCATCAAGGAGTTTCATCGCTTCAAGTGGCGATGGAGGCA 4101

QY 62 CCGTGAAACGGCCACGAGTTTCGAGATCGAGGGCGAGGGCGCGCCCTTCCGCTGGGACATCC 121
 Db CCGTGAAACGGCCACGAGTTTCGAGATCGAGGGCGAGGGCGCGCCCTTCCGCTGGGACATCC 4161

QY 122 ACAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCCGCTGGGACATCC 181
 Db ACAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCCGCTGGGACATCC 4221

QY 182 TGTCCCCCAGTTCAGTACGGCTTCAAGGTGACGTGAAGCACCCCGCGACATCCCGG 241
 Db TGTCCCCCAGTTCAGTACGGCTTCAAGGTGACGTGAAGCACCCCGCGACATCCCGG 4281

QY 242 ACTACAGAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATTCGAGG 301
 Db ACTACAGAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATTCGAGG 4341

QY 302 ACGGGCGGTGGCGACCGTGACCCAGGACTCTCTCCCTGAGGACGGCTTTCATCTACA 361
 Db ACGGGCGGTGGCGACCGTGACCCAGGACTCTCTCCCTGAGGACGGCTTTCATCTACA 4401

QY 362 AGGTGAAGTTTCATCGCGTGAACTTCCCTCCGACGGCCCGTGTATGAGAGAGACCA 421
 Db AGGTGAAGTTTCATCGCGTGAACTTCCCTCCGACGGCCCGTGTATGAGAGAGACCA 4461

QY 422 TGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGGCGTGTGAAGGGCGAGA 481
 Db TGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGGCGTGTGAAGGGCGAGA 4521

QY 482 TCCACAAGCCCTGAAGCTGAAGGACGGCGGCACTACTCTGGTGGAGTTCAAGTCCATCT 541
 Db TCCACAAGCCCTGAAGCTGAAGGACGGCGGCACTACTCTGGTGGAGTTCAAGTCCATCT 4581

QY 542 ACATGGCCAGAGCCGCTGAGCTGCGCGGTACTACTACTGACACACCGAGCTGGACA 601
 Db ACATGGCCAGAGCCGCTGAGCTGCGCGGTACTACTACTGACACACCGAGCTGGACA 4641

QY 602 TCACCTCCCAACAGGAGTACTACCATCGTGGAGCAGTACAGCGCCACCGAGGCGCGCC 661
 Db TCACCTCCCAACAGGAGTACTACCATCGTGGAGCAGTACAGCGCCACCGAGGCGCGCC 4701

QY 662 ACCACTGTTCCTGTA 677
 Db ACCACTGTTCCTGTA 4717

RESULT 13
 AAD10003
 ID AAD10003 standard; DNA; 5436 BP.
 XX
 AC AAD10003;
 XX
 DT 12-SBP-2001 (first entry)
 XX
 DE Plasmid pSPW6 to express goosecoid cDNA fused to red fluorescent protein.
 XX
 KW Mouse; fertility; reproduction; gametogenesis; microinjection; CMV;
 KW cytomegalovirus; SV40; simian virus 40; infection; goosecoid gene;
 KW plasmid pSPW 6; embryogenesis; ds.

OS Mus musculus.
 OS Cytomegalovirus.
 OS Rhesus macaque polyoma virus.
 OS Unidentified.
 OS Chimeric.
 XX
 FN WO200148224-A1.
 XX
 PD 05-JUL-2001.
 XX
 PF 22-DEC-2000; 2000WO-AU001596.
 XX
 PR 24-DEC-1999; 99AU-00004884.
 XX
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 XX
 PI Thresher R, Hinds L, Hardy C, Whyard S, Vignarajan S, Grewe PM;
 PI Patil J;
 XX
 DR WPI; 2001-425672/45.
 XX
 XX Novel construct for preventing embryogenesis in animals comprises native
 PT promoter, blocking DNA which abrogates function of crucial gene and
 PT Genetic switch to regulate expression/repression of blocker/gene
 PT knockout.
 XX
 PS Example 13; Page 215-217; 241pp; English.
 XX
 CC The invention relates to a construct which allows animals to be bred in
 CC captivity but renders them infertile in the wild by allowing reversible
 CC control over fertility and reproduction. The construct comprises a native
 CC promoter, a blocking DNA sequence contoured for and designed to abrogate
 CC a crucial gene's function or to cause its mis-expression, and a genetic
 CC switch to regulate controlled expression/repression of the blocker/gene
 CC knockout. The construct is useful for preventing embryogenesis or
 CC gametogenesis in animals by stably transforming an animal cell with the
 CC construct by microinjection, transfection or infection, where the
 CC construct stably integrates into the genome by homologous recombination,
 CC and implanting the cell into a host organism, where a whole animal
 CC develops from the implanted cell. The present sequence is plasmid pSPW 6
 CC containing mouse goosecoid promoter expressing goosecoid cDNA fused to
 CC red fluorescent protein. The plasmid contains CMV promoter and SV40 polyA
 CC and SV40 ori
 XX
 SQ Sequence 5436 BP; 1247 A; 1525 C; 1492 G; 1172 T; 0 U; 0 Other;

Query Match 97.3%; Score 660; DB 4; Length 5436;
 Best Local Similarity 98.5%; Pred. No. 2.2e-106;
 Matches 666; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 TGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATGGCGTTTCAAGGTGCGCATGGAGGCA 61
 Db TGGCTCTCTCCAGAACGTCATCAAGGAGTTTCATGGCGTTTCAAGGTGCGCATGGAGGCA 1486

QY 62 CCGTGAAACGGCCACGAGTTTCGAGATCGAGGGCGAGGGCGCGCCCTTACGAGGGCC 121
 Db CCGTGAAACGGCCACGAGTTTCGAGATCGAGGGCGAGGGCGCGCCCTTACGAGGGCC 1546

QY 122 ACAACACCGTGAAGTTTCGAGGTGACCAAGGGCGGCCCTTCCGCTGGGACATCC 181
 Db ACAACACCGTGAAGTTTCGAGGTGACCAAGGGCGGCCCTTCCGCTGGGACATCC 1606

QY 182 TGTCCCCCAGTTCAGTACGGCTTCAAGGTGACGTGAAGCACCCCGCGACATCCCG 241
 Db TGTCCCCCAGTTCAGTACGGCTTCAAGGTGACGTGAAGCACCCCGCGACATCCCG 1666

QY 242 ACTACAGAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATCAATTCGAGG 301
 Db ACTACAGAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAATCAATTCGAGG 1726

QY 302 ACGGCGCGGTGGCGACCGGTGACCCAGGACTCTCTCCCTTGCAGGACGGTGTCTTCTACA 361
 Db ACGGCGCGGTGGGTGACCGGTGACCCAGGACTCTCTCCCTTGCAGGACGGTGTCTTCTACA 1786

QY 362 AGGTGAAGTTTCATCGGCTGAAGTTCCTCCGACGGCCCGTGTGATGCAAGAAGACCA 421
Db 1787 AGGTGAAGTTTCATCGGCTGAAGTTCCTCCGACGGCCCGTGTGATGCAAGAAGACCA 1846
QY 422 TGGGCTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCTGTCTGAAGGGCGAGA 481
Db 1847 TGGGCTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCTGTCTGAAGGGCGAGA 1906
QY 482 TCACAAAGCCCTGAAGCTGAAGGAGCGGCGCACTACTCTGTGAGTTCAAGTCCATCT 541
Db 1907 TCACAAAGCCCTGAAGCTGAAGGAGCGGCGCACTACTCTGTGAGTTCAAGTCCATCT 1966
QY 542 ACATGGCAAGAGCCCGTGCAGTCCCGGCTACTACTACGTGGACACCAAGCTGGACA 601
Db 1967 ACATGGCAAGAGCCCGTGCAGTCCCGGCTACTACTACGTGGACTCCAAGCTGGACA 2026
QY 602 TCACCTCCCAAGAGGACTACACCATGTGTGAGCAGTACGAGCGACCGAGGGCGGC 661
Db 2027 TCACCTCCCAAGAGGACTACACCATGTGTGAGCAGTACGAGCGACCGAGGGCGGC 2086
QY 662 ACCACCTGTTCTCTGTA 677
Db 2087 ACCACCTGTTCTCTGTA 2102

RESULT 14
ID ADE24111 standard; DNA; 6893 BP.
XX AC ADE24111;
XX DT 29-JAN-2004 (first entry)
XX DE Proviral plasmid DNA SFR.

XX site-specific DNA recombination; Cre recombinase;
XX protein-transduction domain; gene activation; gene inactivation;
XX chromosomal translocation; reporter protein; ds; circular.
XX Synthetic.

Key Location/Qualifiers
FT misc_RNA 1..435
FT repeat_region /*tag= a
/*note= "plasmid pUC fragment"
436..996
/*tag= b
/*standard_name= "LTR"
/*note= "MPSV 5'-LTR (Delta-31)"
997..1014
FT primer_bind /*tag= c
/*note= "Unidentified primer"
1015..1560
FT 5'UTR /*tag= d
1561..1599
FT misc_RNA /*tag= e
/*note= "loxP1 sequence"
1600..2289
FT CDS /*tag= f
/*product= "DsRed1"
/*note= "from Discosoma sp"
2290..2343
FT misc_RNA /*tag= g
/*note= "loxP2 sequence"
2347..3150
FT CDS /*tag= h
/*product= "egfp"
/*note= "from Aequorea victoria"
3164..4019
FT CDS /*tag= i
/*product= "wPRE"
4020..4080
FT 3'UTR

FT repeat_region /*tag= j
4081..4641
FT /*tag= k
/*standard_name= "LTR"
FT /*note= "3'-LTR"
4642..6893
FT /*tag= l
/*note= "plasmid pUC fragment"
XX WO2003070931-A2.
XX 28-AUG-2003.
XX 19-FEB-2003; 2003WO-EP001680.
XX 21-FEB-2002; 2002DE-01007313.
PR 16-JUL-2002; 2002DE-01032196.
XX (VISI-) VISION 7 GMBH.
XX Baum C, Will E, Ostertag W, Klump H, Schiedlmeier B;
PI WPI; 2003-767353/72.
XX In vitro or in vivo site-specific DNA recombination, useful e.g. for gene
inactivation, using Cre recombinase that lacks heterologous protein-
transduction domain.
XX Claim 23; SEQ ID NO 8; 84pp; German.
XX This invention describes a novel method of site-specific DNA
recombination in eukaryotic cells in vitro comprising using a
bacteriophage P1 Cre recombinase that lacks heterologous protein-
transduction domains. The invention also contains a reporter system for
detecting site-specific DNA recombination in eukaryotic cells.
XX Recombination with Cre recombinase lacking heterologous protein-
transduction domains is used for therapeutic site-specific recombination
in eukaryotic cells, in vivo or in vitro, e.g. for partial inactivation
or activation of genes, also to create a chromosomal translocation. The
modified recombinase can enter intact (not electrically or chemically
treated) cells when added to culture medium and catalyzes recombination
events in the nucleus with recombination rate over 50%, after a single
application. The Cre-recombinase does not need to be supplied from
nucleic acid introduced into the cell, so its effect is immediate but
short-lived, limiting cytotoxic effects; eliminating the risk of non-
specific integration of DNA into the genome and providing quick results.
XX This sequence represents the proviral plasmid Cre reporter construct SFR
used in the method described in the invention.

XX Sequence 6893 BP; 1564 A; 1940 C; 1778 G; 1611 T; 0 U; 0 Other;
Query Match 97.3%; Score 659.6; DB 9; Length 6893;
Best Local Similarity 98.7%; Pred. No. 2.5e-106;
Matches 665; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGGCTCTCTCCGAGAACGTCTATCCAGTTCATCGCTTCAAGTGGCATGGAGGCA 61
Db 1604 TCGCTCTCTCCGAGAACGTCTATCCAGTTCATCGCTTCAAGTGGCATGGAGGCA 1663
QY 62 CCGTGAACGGCCACGAGTTCGAGATCGAGGCGGAGGGCGGCCCTTACGAGGGCC 121
Db 1664 CCGTGAACGGCCACGAGTTCGAGATCGAGGCGGAGGGCGGCCCTTACGAGGGCC 1723
QY 122 ACAACACCGTGAAGTTGAAGTGAACGAGGGCGGCCCTTCCCTTGGGACATCC 181
Db 1724 ACAACACCGTGAAGTGAAGTGAACGAGGGCGGCCCTTCCCTTGGGACATCC 1783
QY 182 TGTCCCCCAGTTCAGTACCGCTTCCAGGTGTAGTGAAGCAGCCCGGACATCCCG 241
Db 1784 TGTCCCCCAGTTCAGTACCGCTTCCAGGTGTAGTGAAGCAGCCCGGACATCCCG 1843
QY 242 ACTACAAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGCGGTGATGAATCCAGG 301

Db 605 TCACCTCCCAACAGGAGTACACCATCGTGGAGCAGTACGAGCGCAGCGCGCC 664
QY 662 ACCACCTGTTCTCTGTA 677
Db 665 ACCACCTGTTCTCTGGA 680

Search completed: August 3, 2004, 11:29:37
Job time : 453.858 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: August 3, 2004, 10:03:29 ; Search time 4535.5 Seconds
(without alignments)
6479.234 Million cell updates/sec
Title: US-10-081-864-14
Perfect score: 678
Sequence: 1 atggctctctccgagaagct.....gcacaccactgttctctgtaa 678
Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0
Searched: 3470272 seqs, 21871516995 residues
Total number of hits satisfying chosen parameters: 6940544
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	678	100.0	678	6	AX824732	AX824732 Sequence
2	676.4	99.8	678	6	AX370408	AX370408 Sequence
3	672.2	99.1	1050	6	AX666133	AX666133 Sequence
4	670.2	98.8	675	6	AX824731	AX824731 Sequence
5	668.4	98.6	678	6	AX370406	AX370406 Sequence
6	665.2	98.1	678	6	AX370404	AX370404 Sequence
7	665.2	98.1	678	6	AX824725	AX824725 Sequence
8	661.6	97.6	4692	6	AX463702	AX463702 Sequence
9	661.6	97.6	9320	6	AX663075	AX663075 Sequence
10	659.6	97.3	6893	6	AX823860	AX823860 Sequence
11	649.8	95.8	10141	12	AY342347	AY342347 Red H-Pel
12	649.8	95.8	10276	12	AY342348	AY342348 Red H-Sti
13	625.6	92.9	681	12	AF506025	AF506025 Synthetic
14	627.6	92.6	1395	12	AF506026	AF506026 Synthetic
15	614.2	90.6	681	6	AX233584	AX233584 Sequence
16	614.2	90.6	713	6	AX233627	AX233627 Sequence
17	568.6	83.9	678	12	AF506027	AF506027 Synthetic
18	556.4	82.1	678	6	AX824729	AX824729 Sequence
19	438	64.6	678	6	AR183915	AR183915 Sequence
20	438	64.6	678	6	AX172854	AX172854 Sequence
21	438	64.6	678	6	AX207715	AX207715 Sequence
22	438	64.6	678	6	AX233581	AX233581 Sequence
23	438	64.6	859	3	AF168419	AF168419 Discosoma
24	438	64.6	859	6	AX463698	AX463698 Sequence
25	432	63.7	666	6	AX348043	AX348043 Sequence
26	432	63.7	666	6	AX353910	AX353910 Sequence
27	422	62.2	898	6	AX686888	AX686888 Sequence
28	377.8	55.7	876	3	AF272711	AF272711 Discosoma
29	377.8	55.7	876	6	AX686894	AX686894 Sequence
30	313.4	46.2	707	6	AX824739	AX824739 Sequence
31	308.6	45.5	699	12	AY233273	AY233273 Synthetic
32	305	45.0	4707	12	AY218848	AY218848 Monster G
33	297.2	43.8	654	6	AX824741	AX824741 Sequence
34	296.6	43.7	669	6	AX699925	AX699925 Sequence
35	295	43.5	669	6	AX699921	AX699921 Sequence
36	295	43.5	669	6	AX699927	AX699927 Sequence
37	293.4	43.3	660	6	AX699761	AX699761 Sequence
38	293.4	43.3	693	6	AX699793	AX699793 Sequence
39	292.4	43.1	660	6	AX699825	AX699825 Sequence
40	292.4	43.1	687	6	AX527900	AX527900 Sequence
41	292.4	43.1	1396	6	AX527902	AX527902 Sequence
42	292.4	43.1	1424	6	AX527904	AX527904 Sequence
43	291.8	43.0	660	6	AX699785	AX699785 Sequence
44	291.8	43.0	660	6	AX699809	AX699809 Sequence
45	291.8	43.0	663	6	AX699755	AX699755 Sequence

ALIGNMENTS

RESULT 1
AX824732
LOCUS AX824732 678 bp DNA linear PAT 11-DEC-2003
DEFINITION Sequence 14 from Patent WO02068459.
ACCESSION AX824732
VERSION AX824732.1 GI:39750595
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Non aggregating fluorescent proteins and methods for using the same
Patent: WO 02068459-A 14 06-SEP-2002;
Location/Qualifiers

REFERENCE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	
-----------	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	--

RESULT 4	linear	PAT 11-DEC-2003
AX824731	675 bp	DNA
LOCUS		
DEFINITION	Sequence 13 from Patent WO02068459.	
AX824731		
ACCESSION		
AX824731.1	GI:39750594	
KEYWORDS		

RESULT 5	linear	PAT 16-FEB-2003
AX370406	678 bp	DNA
LOCUS		
DEFINITION	Sequence 3 from Patent WO0196373.	
AX370406		
ACCESSION		
AX370406		

51

VERSION AX370406.1 GI:18857491
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.

REFERENCE 1
AUTHORS Fradkov, A.F. and Tersikh, A.
TITLE Fluorescent timer proteins and methods for their use
JOURNAL Patent: WO 0196373-A 3 20-DEC-2001;
Clontech Laboratories Inc. (US)
LOCATION/Qualifiers

FEATURES
source 1. 678
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="variant of sequence from Discosoma sp."

ORIGIN

Query Match 98.6%; Score 668.4; DB 6; Length 678;
Best Local Similarity 99.1%; Pred. No. 1e-78;
Matches 672; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTTCATGCGTTCAAGGTGCGCATGGAGGC 60
DB 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTTCATGCGTTCAAGGTGCGCATGGAGGC 60

QY 61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTTACGAGGC 120
DB 61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTTACGAGGC 120

QY 121 CACAAACCGGTGAAGTTGAGGTGACCAAGGGCGGGCGCCCTTCCCTTCCCTTGGGACATC 180
DB 121 CACAAACCGGTGAAGTTGAGGTGACCAAGGGCGGGCGCCCTTCCCTTCCCTTGGGACATC 180

QY 181 CTGTCCCCCAGTTCCAGTACGGCTCCAGGTGACCAAGGGCGGGCGCCCTTCCCTTCCCTTGGGACATC 240
DB 181 CTGTCCCCCAGTTCCAGTACGGCTCCAGGTGACCAAGGGCGGGCGCCCTTCCCTTCCCTTGGGACATC 240

QY 241 GACTACAAAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAATTCGAG 300
DB 241 GACTACAAAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAATTCGAG 300

QY 301 GACGGCGGTGCGACCGGTGACCGAGGACTCTCTCCCTGAGGACCGGCTTCTCTCTAC 360
DB 301 GACGGCGGTGCGACCGGTGACCGAGGACTCTCTCCCTGAGGACCGGCTTCTCTCTAC 360

QY 361 AAGTGAAGTTATCGGCGTGAATCTTCCCTCCGAGCGCCCTGATGAGAGAGAGACC 420
DB 361 AAGTGAAGTTATCGGCGTGAATCTTCCCTCCGAGCGCCCTGATGAGAGAGAGACC 420

QY 421 ATGSGTGGAGGCGCTTCCAGCGGCGCTGTACCCCGGAGCGCGCTGATGAGAGAGAG 480
DB 421 ATGSGTGGAGGCGCTTCCAGCGGCGCTGTACCCCGGAGCGCGCTGATGAGAGAGAG 480

QY 481 ATCCACAAGCCCTGAAGCTGAAGGACGGCGGCGGCTTCTCTCTGAGGTTCAAGTCCATC 540
DB 481 ATCCACAAGCCCTGAAGCTGAAGGACGGCGGCGGCTTCTCTCTGAGGTTCAAGTCCATC 540

QY 541 TACATGGCCCAAGAGCCGCTGACGCTGCCCGGTACTACTAGTGACACCAAGCTGGAC 600
DB 541 TACATGGCCCAAGAGCCGCTGACGCTGCCCGGTACTACTAGTGACACCAAGCTGGAC 600

QY 601 ATCACTTCCCAACAGGAGGACTACACCATCTGTGAGCAGTACGAGCGCACCGAGGCGCGC 660
DB 601 ATCACTTCCCAACAGGAGGACTACACCATCTGTGAGCAGTACGAGCGCACCGAGGCGCGC 660

QY 661 CACCACCTGTCTCTGTAA 678
DB 661 CACCACCTGTCTCTGTAA 678

LOCUS AX370404
DEFINITION Sequence 1 from Patent WO0196373.
ACCESSION AX370404
VERSION AX370404.1 GI:18857490
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.

REFERENCE 1
AUTHORS Fradkov, A.F. and Tersikh, A.
TITLE Fluorescent timer proteins and methods for their use
JOURNAL Patent: WO 0196373-A 1 20-DEC-2001;
Clontech Laboratories Inc. (US)
LOCATION/Qualifiers

FEATURES
source 1. 678
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="variant of sequence from Discosoma sp."

ORIGIN

Query Match 98.1%; Score 665.2; DB 6; Length 678;
Best Local Similarity 98.8%; Pred. No. 2.6e-78;
Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTTCATGCGTTCAAGGTGCGCATGGAGGC 60
DB 1 ATGCGCTCTCCGAGAACGTCATCACCGAGTTTCATGCGTTCAAGGTGCGCATGGAGGC 60

QY 61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTTACGAGGC 120
DB 61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTTACGAGGC 120

QY 121 CACAAACCGGTGAAGTTGAGGTGACCAAGGGCGGGCGCCCTTCCCTTCCCTTGGGACATC 180
DB 121 CACAAACCGGTGAAGTTGAGGTGACCAAGGGCGGGCGCCCTTCCCTTCCCTTGGGACATC 180

QY 181 CTGTCCCCCAGTTCCAGTACGGCTCCAGGTGACCAAGGGCGGGCGCCCTTCCCTTCCCTTGGGACATC 240
DB 181 CTGTCCCCCAGTTCCAGTACGGCTCCAGGTGACCAAGGGCGGGCGCCCTTCCCTTCCCTTGGGACATC 240

QY 241 GACTACAAAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAATTCGAG 300
DB 241 GACTACAAAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAATTCGAG 300

QY 301 GACGGCGGTGCGACCGGTGACCGAGGACTCTCTCCCTGAGGACCGGCTTCTCTCTAC 360
DB 301 GACGGCGGTGCGACCGGTGACCGAGGACTCTCTCCCTGAGGACCGGCTTCTCTCTAC 360

QY 361 AAGTGAAGTTATCGGCGTGAATCTTCCCTCCGAGCGCCCTGATGAGAGAGAGACC 420
DB 361 AAGTGAAGTTATCGGCGTGAATCTTCCCTCCGAGCGCCCTGATGAGAGAGAGACC 420

QY 421 ATGSGTGGAGGCGCTTCCAGCGGCGCTGTACCCCGGAGCGCGCTGATGAGAGAGAG 480
DB 421 ATGSGTGGAGGCGCTTCCAGCGGCGCTGTACCCCGGAGCGCGCTGATGAGAGAGAG 480

QY 481 ATCCACAAGCCCTGAAGCTGAAGGACGGCGGCGGCTTCTCTCTGAGGTTCAAGTCCATC 540
DB 481 ATCCACAAGCCCTGAAGCTGAAGGACGGCGGCGGCTTCTCTCTGAGGTTCAAGTCCATC 540

QY 541 TACATGGCCCAAGAGCCGCTGACGCTGCCCGGTACTACTAGTGACACCAAGCTGGAC 600
DB 541 TACATGGCCCAAGAGCCGCTGACGCTGCCCGGTACTACTAGTGACACCAAGCTGGAC 600

QY 601 ATCACTTCCCAACAGGAGGACTACACCATCTGTGAGCAGTACGAGCGCACCGAGGCGCGC 660
DB 601 ATCACTTCCCAACAGGAGGACTACACCATCTGTGAGCAGTACGAGCGCACCGAGGCGCGC 660

QY 661 CACCACCTGTCTCTGTAA 678
DB 661 CACCACCTGTCTCTGTAA 678

```
RESULT 7
AX824725
LOCUS AX824725 678 bp DNA linear PAT 11-DEC-2003
DEFINITION Sequence 7 from Patent WO02068459.
ACCESSION AX824725
VERSION AX824725.1 GI:39750591
KEYWORDS
SOURCE Discosoma sp.
ORGANISM Discosoma sp.
Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE
1
AUTHORS Non aggregating fluorescent proteins and methods for using the same
TITLE Patent: WO 02068459-A 7 06-SEP-2002;
JOURNAL
FEATURES
source
/organism="Discosoma sp."
/mol_type="unassigned DNA"
/db_xref="taxon:86600"
ORIGIN
Query Match 98.1%; Score 665.2; DB 6; Length 678;
Best Local Similarity 98.8%; Pred. No. 2.6e-78;
Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTCATCGCTTCAAGTGGCGATGGAGGC 60
DB 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTCATCGCTTCAAGTGGCGATGGAGGC 60
QY 61 ACCGTGAACGGCCACAGTTCGAGATCGAGGGCGAGGGCGGGCCGCCCTACGAGGC 120
DB 61 ACCGTGAACGGCCACAGTTCGAGATCGAGGGCGAGGGCGGGCCGCCCTACGAGGC 120
QY 121 CACAAACACCGTGAAGTGAAGTGAACAGGGGGCGCCCTTGCCTTGGGACATC 180
DB 121 CACAAACACCGTGAAGTGAAGTGAACAGGGGGCGCCCTTGCCTTGGGACATC 180
QY 181 CTGTCCCCCAGTTCAGTACGCTCAAGTGTGTAAGTGAACACCGCGACATCCCC 240
DB 181 CTGTCCCCCAGTTCAGTACGCTCAAGTGTGTAAGTGAACACCGCGACATCCCC 240
QY 241 GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGTGATGAACCTTCGAG 300
DB 241 GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGTGATGAACCTTCGAG 300
QY 301 GACGGCGGTGGGACCGTGAACCGAGTCTCCTCGAGCGGCTGCTTCATCTAC 360
DB 301 GACGGCGGTGGGACCGTGAACCGAGTCTCCTCGAGCGGCTGCTTCATCTAC 360
QY 361 AAGTGAAGTTCATCGGCGTGAACCTTCCCTCGAGCGGCTGATGAGAGAGACC 420
DB 361 AAGTGAAGTTCATCGGCGTGAACCTTCCCTCGAGCGGCTGATGAGAGAGACC 420
QY 421 ATGGGTGGAGGCTCCACCGAGCGCTGTACCCCGGAGCGGCTGTGAAGGGGAG 480
DB 421 ATGGGTGGAGGCTCCACCGAGCGCTGTACCCCGGAGCGGCTGTGAAGGGGAG 480
QY 481 ATCCACAGGCGCTGAAGCTGAAGAGCGGCGGCTTCTGTGGAGTTCAGTCCATC 540
DB 481 ATCCACAGGCGCTGAAGCTGAAGAGCGGCGGCTTCTGTGGAGTTCAGTCCATC 540
QY 541 TACATGCCAAGAGCCCGTGCAGCTCCCGGCTTACTACTGTGGACACCAAGCTGGAC 600
DB 541 TACATGCCAAGAGCCCGTGCAGCTCCCGGCTTACTACTGTGGAGTTCAGTCCATC 600
QY 601 ATACCTCCCAACAGGAGTACCATCGTGGAGCAGTACGAGCCACCGAGGGCGGC 660
DB 601 ATACCTCCCAACAGGAGTACCATCGTGGAGCAGTACGAGCCACCGAGGGCGGC 660
QY 661 CACCACCTGTTCCTGTAA 678
```

```
Db 661 CACCACCTGTTCCTGTAA 678
RESULT 8
AX463702
LOCUS AX463702 4692 bp DNA linear PAT 15-JUL-2002
DEFINITION Sequence 16 from Patent WO0248338.
ACCESSION AX463702
VERSION AX463702.1 GI:21886461
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS Lichtenberg-Frat, H.
TITLE Yeast strain for testing the geno- and cytotoxicity of complex
environmental contamination
JOURNAL Patent: WO 0248338-A 16 20-JUN-2002;
FEATURES
source
Location/Qualifiers
1..4692
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="Vektor pDSRed1-N1"
ORIGIN
Query Match 97.6%; Score 661.6; DB 6; Length 4692;
Best Local Similarity 98.7%; Pred. No. 5.1e-78;
Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGGCTCTCTCCGAGAACGTCATCACCGAGTTCATCGCTTCAAGTGGCGATGGAGGC 61
DB 683 TGGCTCTCTCCGAGAACGTCATCACCGAGTTCATCGCTTCAAGTGGCGATGGAGGC 742
QY 62 CCGTGAACGGCCACAGTTCGAGATCGAGGGCGAGGGCGGGCCGCCCTACGAGGCC 121
DB 743 CCGTGAACGGCCACAGTTCGAGATCGAGGGCGAGGGCGGGCCGCCCTACGAGGCC 802
QY 122 ACAAACACCGTGAAGTGAAGTGAACAGGGCGGCGCCCTTGCCTTGGGACATCC 181
DB 803 ACAAACACCGTGAAGTGAAGTGAACAGGGCGGCGCCCTTGCCTTGGGACATCC 862
QY 182 TGTCTCCCGAGTTCAGTACGCTTCCCGAGGGCTTCCCGAGGGCTTCCCGAGGGCTTCCCG 241
DB 863 TGTCTCCCGAGTTCAGTACGCTTCCCGAGGGCTTCCCGAGGGCTTCCCGAGGGCTTCCCG 922
QY 242 ACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGTGATGAATTCGAGG 301
DB 923 ACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGTGATGAATTCGAGG 982
QY 302 ACGGGCGGTGGCGACCGTGAACCGAGTCTTCCCTCGAGGAGCGGTCTCATCTACA 361
DB 983 ACGGGCGGTGGTGAACCGTGAACCGAGTCTTCCCTCGAGGAGCGGTCTCATCTACA 1042
QY 362 AGGTGAAGTTCATCGGCGTGAACCTTCCCGAGGGCTTCCCGAGGGCTTCCCGAGGGCTTCCCG 421
DB 1043 AGGTGAAGTTCATCGGCGTGAACCTTCCCGAGGGCTTCCCGAGGGCTTCCCGAGGGCTTCCCG 1102
QY 422 TGGGTGGAGGCTTCCACCGAGCGCTGTACCCCGGAGCGGTGCTGAAGGGCGAGA 481
DB 1103 TGGGTGGAGGCTTCCACCGAGCGCTGTACCCCGGAGCGGTGCTGAAGGGCGAGA 1162
QY 482 TCCACAAGGCGCTGAAGTGAAGGACGCGGCGCTTCTGTGGAGTTCAGTCCATCT 541
DB 1163 TCCACAAGGCGCTGAAGTGAAGGACGCGGCGCTTCTGTGGAGTTCAGTCCATCT 1222
QY 542 ACATGGCCCAAGAGCCCGTGCAGTTCGCCGGCTTACTACTGTGGACACCAAGCTGGACA 601
DB 1223 ACATGGCCCAAGAGCCCGTGCAGTTCGCCGGCTTACTACTGTGGACTCCAAGCTGGACA 1282
QY 602 TACCTCCCAACAGGAGTACCATCTGTGGAGCTACCGAGCGCATCCGAGGGCGGCC 661
```

Db 1283 TCACCTCCACACAGGAGCTACACCATCGTGGAGCAGTACGAGCGGACCGAGGGCGGCC 1342

QY 662 ACCACCTGTTCTGTGA 677
 |||||
 Db 1343 ACCACCTGTTCTGTGA 1358

RESULT 9
 AX663075
 LOCUS AX663075
 DEFINITION Sequence 20 from Patent WO02070740. linear PAT 24-MAR-2003
 ACCESSION AX663075
 VERSION AX663075.1 GI:29169369
 KEYWORDS synthetic construct
 SOURCE synthetic construct
 ORGANISM artificial sequences.

REFERENCE 1
 Wiesmuller L.
 Test system for determining gene toxicities
 Patent: WO 02070740-A 20 12-SEP-2002;
 Wiesmuller, Lisa (DE)
 JOURNAL Location/Qualifiers
 FEATURES
 source 1..9320
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Plasmid p5-Puro-CMV- (N'-EGFP)-CMV-Red- (EGFP-EJ)"
 misc_feature 1..1552
 /note="Retroviral Vektor p5NM"
 1617..2216
 /note="unnamed protein product; Puromycin-Resistenzgen aus pRetroon (Clontech, Palo Alto, CA, USA)"
 CDS
 /codon_start=1
 /transl_table=11
 /protein_id="CAD80174.1"
 /db_xref="GI:29169370"
 /db_xref="REMBL:CAD80174"
 /translations="MTEYKPTVRLATDDVPRVRLTAAAFADYPATRTHTVDPDRHIE RVTELOELFTRVLDIGKVVADGAAVAVTTTPSVEAGAVFAIGPRMAELGSR LAACQOMEGILLAPHRKEPAKELATVGVSPDHOGKGLGSAVLPVGEAERAGVPAFL ETAPRNLPYERLGTIVTADVECPKDRATWCTRPGA"
 2267..2848
 /note="CMV-Promotor aus pEGFP-N1 (Clontech, Palo Alto, CA, USA)"
 promoter 2906..>3349
 /note="unnamed protein product; N'-EGFP, abgeleitet von EGFP aus pEGFP-N1 (Clontech, Palo Alto, CA, USA)"
 CDS
 /codon_start=1
 /transl_table=11
 /protein_id="CAD80175.1"
 /db_xref="GI:29169371"
 /db_xref="REMBL:CAD80175"
 /translations="MYSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGSDATYKGLT LKICTITGLPVPWPLVTLTVGVOCFSRYDPHMKQHDFFKSNMPEGVQVTRTIFPK DCGNYKTRAEVKEGTLVNRILKIDPKEDGNILGHKLEYNR"
 3374..3392
 /note="Retroviral Vektor p5NM"
 3411..3992
 /note="CMV-Promotor aus pbsRed1-N1 (Clontech, Palo Alto, CA, USA)"
 misc_feature 4038..4718
 /note="unnamed protein product; Red aus pbsRed1-N1 (Clontech, Palo Alto, CA, USA)"
 promoter
 /codon_start=1
 /transl_table=11
 /protein_id="CAD80176.1"
 /db_xref="GI:29169372"
 /db_xref="REMBL:CAD80176"
 /translations="MTRSSKNVTKEMRFKVRMEGTVNGHFEIISGEGRPYEGHNT VKLVTKGGLPFAWDILSPQYGSKVYKRPADIPDYKLSFPEGFKEWVMFED GGVTVTQSSLDQGCFFIVKXFIQVNFPSDGFVMQKTMWEASTERLYPRDGLVKG EITHALKLKDGGHYLVEFKSIYNAKKPVQLPGYYVDSKLDITSHEDYITVEQVERT

misc_feature EGRHHLFL"
 4766..5509
 /note="EGFP-EJ, abgeleitet von EGFP aus pEGFP-N1 (Clontech, Palo Alto, CA, USA)"
 misc_feature 5527..9320
 /note="Retroviral Vektor p5NM"

ORIGIN

Query Match 97.6%; Score 661.6; DB 6; Length 9320;
 Best Local Similarity 98.7%; Pred. No. 4.4e-78;
 Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGCTCTCTCCGAGAACGTTCATCCAGGATTCAGGTTTCATCGGTTCAAGTGGCATGGAGGCA 61
 |||||
 Db 4042 TGGCTCTCTCCAGAACGTTCATCAAGGAGTTTCATCGGTTCAAGTGGCATGGAGGCA 4101
 |||||
 QY 62 CCGTGAACGGCCACGAGTTCGAGATCGAGGCGAGGGCGAGGGCGGCGCCCTACGAGGGCC 121
 |||||
 Db 4102 CCGTGAACGGCCACGAGTTCGAGATCGAGGCGAGGGCGAGGGCGGCGCCCTACGAGGGCC 4161
 |||||
 QY 122 ACAACACCGTGAAGTTGAAGTGACCAAGGGGGGCGCCCTGCCCTTCGCTGGGACATCC 181
 |||||
 Db 4162 ACAACACCGTGAAGTTGAAGTGACCAAGGGGGGCGCCCTGCCCTTCGCTGGGACATCC 4221
 |||||
 QY 182 TGTCCCGCCAGTTCACGATACGCTTCCAAAGTGTAGCTCAAGCAAGCGGCGGCGGCGGCGGCGG 241
 |||||
 Db 4222 TGTCCCGCCAGTTCACGATACGCTTCCAAAGTGTAGCTCAAGCAAGCGGCGGCGGCGGCGG 4281
 |||||
 QY 242 ACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGCTGATGAATTCGAGG 301
 |||||
 Db 4282 ACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGCTGATGAATTCGAGG 4341
 |||||
 QY 302 ACGGGCGGTGGGACCGTGAACCGGACTCTCTCCCTGCAGGCGGCTTCTCATCTACA 361
 |||||
 Db 4342 ACGGGCGGTGGTGAACCGGACTCTCTCCCTGCAGGCGGCTTCTCATCTACA 4401
 |||||
 QY 362 AGGTGAAGTTCATCGGCGTGAATTCCTCCGAGCGGCGGCTGATGAGAAGAACCA 421
 |||||
 Db 4402 AGGTGAAGTTCATCGGCGTGAATTCCTCCGAGCGGCGGCTGATGAGAAGAACCA 4461
 |||||
 QY 422 TGGGTGGAGGCTTCCACCGAGCGCTGTACCCCGGCGGCGGCTGCTGAGGCGGAGA 481
 |||||
 Db 4462 TGGGTGGAGGCTTCCACCGAGCGCTGTACCCCGGCGGCGGCTGCTGAGGCGGAGA 4521
 |||||
 QY 482 TCCACAGGCGCTTCAAGCTGAAGCGGCGGCGGCGGCGGCGGCTTCAAGTCCATCT 541
 |||||
 Db 4522 TCCACAGGCGCTTCAAGCTGAAGCGGCGGCGGCGGCGGCGGCTTCAAGTCCATCT 4581
 |||||
 QY 542 ACATGGCCAGAAGCGGCTGAGTCCCGGCGGCTACTACTAGTGGACACCAAGCTGACA 601
 |||||
 Db 4582 ACATGGCCAGAAGCGGCTGAGTCCCGGCGGCTACTACTAGTGGACACCAAGCTGACA 4641
 |||||
 QY 602 TCACCTCCACAAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGGCC 661
 |||||
 Db 4642 TCACCTCCACAAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGGCC 4701
 |||||
 QY 662 ACCACCTGTTCTGTGA 677
 |||||
 Db 4702 ACCACCTGTTCTGTGA 4717
 |||||

RESULT 10
 AX823860
 LOCUS AX823860
 DEFINITION Sequence 8 from Patent WO03070931. linear PAT 11-DEC-2003
 ACCESSION AX823860
 VERSION AX823860.1 GI:39750176
 KEYWORDS synthetic construct
 SOURCE synthetic construct
 ORGANISM artificial sequences.
 REFERENCE 1
 AUTHORS Baum, C., Will, E., Ostertag, W., Klump, H. and Schiedlmeier, B.


```
TITLE      Methods for conducting site-specific dna recombination
JOURNAL    Patent: WO 03070931-A 8 28-AUG-2003;
VISION     Vision 7 GmbH (DE)
FEATURES   Location/Qualifiers
            source
            1..6893
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="Beschreibung der kuenstlichen Sequenz: provirale
                Plasmid-DNA; retrovirales Cre-Reporterkonstrukt #Spr#"
            misc_feature
            1..435
                /note="Plasmid-Rueckgrat (pUC)"
            misc_feature
            436..996
                /note="MPSV 5'-LTR (Delta1-31)"
            primer_bind
            997..1014
                /note="PBS (primer binding site)"
            5'UTR
            1015..1560
                /note="1560"
            misc_feature
            1561..1599
                /note="#loxP1#-Sequenz"
            misc_feature
            1600..2289
                /note="DsRed1-Gen aus Discosoma sp."
            misc_feature
            2290..2343
                /note="#loxP2#-Sequenz"
            misc_feature
            2347..3150
                /note="edFP (tag)-Gen aus Aequorea victoria"
            misc_feature
            3154..4019
                /note="#wPRE#"
            3'UTR
            4020..4080
                /note="4080"
            misc_feature
            4081..4641
                /note="3'-LTR"
            misc_feature
            4642..6893
                /note="Plasmid-Rueckgrat (pUC)"

ORIGIN
Query Match      97.3%; Score 659.6; DB 6; Length 6893;
Best Local Similarity 98.7%; Pred. No. 8 5e-78;
Matches 565; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      2 TGGCCTCTCCGAGAACGTCATACCGAGTTTCATGGCTTCAAGGTGCGCATGGAGGCA 61
DB      1604 TGGCCTCTCCGAGAACGTCATACCGAGTTTCATGGCTTCAAGGTGCGCATGGAGGCA 1663

QY      62 CCGTGAACCGCCAGGTTCCAGATCGAGGCGGAGGCGCCGCCCTACGAGGCC 121
DB      1664 CCGTGAACCGCCAGGTTCCAGATCGAGGCGGAGGCGCCGCCCTACGAGGCC 1723

QY      122 ACAACACCGTGAAGTTCAAGGTGACCAAGGCGGCGCCCTCGCCCTTGCCTGGGACATCC 181
DB      1724 ACAACACCGTGAAGTTCAAGGTGACCAAGGCGGCGCCCTCGCCCTTGCCTGGGACATCC 1783

QY      182 TGTCCCCCGAGTTCCAGTACGGCTCCAAAGGTGTAAGTGAAGCACCAGCCGCGATCCCCG 241
DB      1784 TGTCCCCCGAGTTCCAGTACGGCTCCAAAGGTGTAAGTGAAGCACCAGCCGCGATCCCCG 1843

QY      242 ACTACAAGAGCTGCTCTTCCCGAGGCTTCAAGTGGGAGCGCGTGATGAATTCGAGG 301
DB      1844 ACTACAAGAGCTGCTCTTCCCGAGGCTTCAAGTGGGAGCGCGTGATGAATTCGAGG 1903

QY      302 ACGGCGGCGTGGACCGTACCCAGAGTCTCTCCCTGCGAGGCGGTGCTTCAATCTACA 361
DB      1904 ACGGCGGCGTGGTGAACCGTACCCAGAGTCTCTCCCTGCGAGGCGGTGCTTCAATCTACA 1963

QY      362 AGGTGAAGTTTATCGGCGTGAATCTCCCTCCGAGCGCCCGTGCATGCAGAAGAGACCA 421
DB      1964 AGGTGAAGTTTATCGGCGTGAATCTCCCTCCGAGCGCCCGTGCATGCAGAAGAGACCA 2023

QY      422 TGGGCTGGAGGCTCTCCACCGAGGCGCTGTACCCCGCGAGCGGCGTCTGAAGGGCGAGA 481
DB      2024 TGGGCTGGAGGCTCTCCACCGAGGCGCTGTACCCCGCGAGCGGCGTCTGAAGGGCGAGA 2083

QY      482 TCCACAGGCGCTGAAGCTGAAGACGCGCGCCACTACCTGGTGGAGTTCAAGTCCATCT 541
DB      2084 TCCACAGGCGCTGAAGCTGAAGACGCGCGCCACTACCTGGTGGAGTTCAAGTCCATCT 2143
```

```
QY      542 ACATGGCCAAAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGGACACCAAGCTGGACA 601
DB      2144 ACATGGCCAAAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGGACTCCAAAGCTGGACA 2203

QY      602 TCACCTCCCAACAGGAGACTACACCATCTGGAGAGTACGAGCGACCGAGCGCGCCGCC 661
DB      2204 TCACCTCCCAACAGGAGACTACACCATCTGGAGAGTACGAGCGACCGAGCGCGCGCCGCC 2263

QY      662 ACCACCTGTTCCCTG 675
DB      2264 ACCACCTGTTCCCTG 2277

RESULT 11
LOCUS      AY342347              10141 bp      DNA      linear      SYN 21-OCT-2003
DEFINITION Red H-Pelican DsRed.T4 transformation vector, complete sequence.
ACCESSION  AY342347
VERSION    AY342347.1 GI:33358309
KEYWORDS   Red H-Pelican DsRed.T4 transformation vector
SOURCE     Red H-Pelican DsRed.T4 transformation vector
           artificial sequences; vectors.
ORGANISM   Barolo, S. and Posakony, J.W.
REFERENCE  1 (bases 1 to 10141)
           Barolo, S. and Posakony, J.W.
           Advances in Drosophila Transgenic Reporters I: Fast-Maturing RFP
           [DsRed.T4] and Nuclear RFP in Insulated Vectors
           Unpublished
REFERENCE  2 (bases 1 to 10141)
           Barolo, S. and Posakony, J.W.
           Direct Submission
           Submitted (14-JUL-2003) Division of Biology/CDB, UCSD, 4121 Bonner
           Hall, MC 0349, La Jolla, CA 92093-0349, USA
           Location/Qualifiers
           1..10141
           /organism="Red H-Pelican DsRed.T4 transformation vector"
           /mol_type="other DNA"
           /db_xref="taxon:240544"

FEATURES   source
ORIGIN
Query Match      95.8%; Score 649.8; DB 12; Length 10141;
Best Local Similarity 97.5%; Pred. No. 1.5e-76;
Matches 660; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      1 ATGGCTCTCTCCGAGAACGTCATACCGAGTTTCATGGCTTCAAGGTGCGCATGGAGGC 60
DB      285 ATGGCTCTCTCCGAGAACGTCATACCGAGTTTCATGGCTTCAAGGTGCGCATGGAGGC 344

QY      61 ACCGTGAACCGCCAGGTTCCAGATCGAGGCGGAGGCGCGGCGCCCTTACGAGGC 120
DB      345 TCGGTGAACCGCCAGGTTCCAGATCGAGGCGGAGGCGCGGCGCCCTTACGAGGC 404

QY      121 CACACACCGTGAAGTTGAAGTGACCAAGGCGGCGCCCTTCCCTTCCGCTGGGACATC 180
DB      405 ACCCAGACCGCCAAAGTGAAGTGACCAAGGCGGCGCCCTTCCCTTCCGCTGGGACATC 464

QY      181 CTGTCCCGCCAGTTCAGTACCGCTTCCAGGTGTAAGTGAAGCACCAGCGGACATCCC 240
DB      465 CTGTCCCGCCAGTTCAGTACCGCTTCCAGGTGTAAGTGAAGCACCAGCGGACATCCC 524

QY      241 GACTACAAGAAGCTGCTTCCCGAGGCTTCAAGTGGAGCGCGGTGATGAATTCGAG 300
DB      525 GACTACAAGAAGCTGCTTCCCGAGGCTTCAAGTGGAGCGCGGTGATGAATTCGAG 584

QY      301 GACGCGGCGTGGGACCGTGCACCGAGTCTCTCCCTGCGAGGCGGTGCTTCTATCTAC 360
DB      585 GACGCGGCGTGGTGAACCGTGCACCGAGTCTCTCCCTGCGAGGCGGTGCTTCTATCTAC 644

QY      361 AAGGTGAAGTTTCATCGGCGTGAATTCCTCCCTCCGAGCGCCCGTGTATGAGAAGAGACC 420
DB      645 AAGGTGAAGTTTCATCGGCGTGAATTCCTCCCTCCGAGCGCCCGTGTATGAGAAGAGACC 704
```

QY 421 ATGGCTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCGTCTCAAGGGCGAG 480
 Db 705 ATGGGCTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCGTCTCAAGGGCGAG 764
 QY 481 ATCCACAAGGCCCTCAAGCTGAAGACGGCGGCGCACTACTCTGGTGGAGTTCAAGTCCATC 540
 Db 765 ATCCACAAGGCCCTCAAGCTGAAGACGGCGGCGCACTACTCTGGTGGAGTTCAAGTCCATC 824
 QY 541 TATATGGCCCAAGAACCGCTGCAGCTGCCCGGCTACTACTACGTGGACCAACAGCTGGAC 600
 Db 825 TATATGGCCCAAGAACCGCTGCAGCTGCCCGGCTACTACTACGTGGAGTCCAACTGGAC 884
 QY 601 ATCACTCCCAACAGAGACTACACCATCTGTGAGAGTAGAGAGCGACCGAGGGCGC 660
 Db 885 ATCACTCCCAACAGAGACTACACCATCTGTGAGAGTAGAGAGCGACCGAGGGCGC 944
 QY 661 CACCACCTGTTCTCTGTA 677
 Db 945 CACCACCTGTTCTCTGTA 961

RESULT 12
 AY342348
 LOCUS
 DEFINITION Red H-Stinger DsRed.T4-NLS transformation vector, complete
 ACCESSION AY342348
 VERSION AY342348.1 GI:33359310
 KEYWORDS
 SOURCE Red H-Stinger DsRed.T4-NLS transformation vector
 ORGANISM
 REFERENCE
 AUTHORS Barolo, S. and Posakony, J.W.
 TITLE Advances in Drosophila Transgenic Reporters I: Fast-Maturing RFP [DsRed.T4] and Nuclear RFP in Insulated Vectors
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 10276)
 AUTHORS Barolo, S. and Posakony, J.W.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUL-2003) Division of Biology/CDB, UCSD, 4121 Bonner Hall, MC 0349, La Jolla, CA 92093-0349, USA
 FEATURES
 source
 1. 10276
 /organism="Red H-Stinger DsRed.T4-NLS transformation vector"
 /mol_type="other DNA"
 /db_xref="taxon:240545"

ORIGIN
 Query Match 95.8%; Score 649.8; DB 12; Length 10276;
 Best Local Similarity 97.5%; Pred. No. 1.5e-76;
 Matches 660; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 ATGGCTCTCCGAGACGTCATCACCGAGTTGATCGCTTCAAGTGGCGATGGAGGC 60
 Db 288 ATGGCTCTCCGAGACGTCATCACCGAGTTGATCGCTTCAAGTGGCGATGGAGGC 347
 QY 61 ACCGTGAACGCCACGAGTTGAGATCGAGAGCGGCGAGGCGCGCCCTACGAGGC 120
 Db 348 TCCGTGAACGCCACGAGTTGAGATCGAGAGCGGCGAGGCGCGCCCTACGAGGC 407
 QY 121 CACAACACCGTGAAGTTGAAGGTGACCAAGCGGCGCCCTTCCCTTCCCTGGGACATC 180
 Db 408 ACCAGACCGCAAGCTGAAGGTGACCAAGCGGCGCCCTTCCCTTCCCTGGGACATC 467
 QY 181 CTGTCCCCCAGTCCAGTACGCTCCAGGTGATCGTGAAGCACCCCGCGACATCCCC 240
 Db 468 CTGTCCCCCAGTCCAGTACGCTCCAGGTGATCGTGAAGCACCCCGCGACATCCCC 527
 QY 241 GACTACAAGAGCTGCTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAACCTTCGAG 300
 Db 528 GACTACAAGAGCTGCTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAACCTTCGAG 587

QY 301 GACGGCGGCTGGCGACCGTGAACCGAGTCTCTCCCTGCAGACGGTGTCTTCAATCTAC 360
 Db 588 GACGGCGGCTGGTGAACCGTGAACCGAGTCTCTCCCTGCAGACGGTGTCTTCAATCTAC 647
 QY 361 AAGGTGAATTCATCGCGTGAATTCCTCCCTCCGACGGCCCGTGTATGAGAAGAGACC 420
 Db 648 AAGGTGAATTCATCGCGTGAATTCCTCCCTCCGACGGCCCGTGTATGAGAAGAGACT 707
 QY 421 ATGGGCTGGAGCGCTCCACCGAGCGCTGTACCCCGCGAGCGGTGTGAGGGCGAG 480
 Db 708 ATGGGCTGGAGCGCTCCACCGAGCGCTGTACCCCGCGAGCGGTGTGAGGGCGAG 767
 QY 481 ATCCACAAGGCCCTGAAGCTGAAGACGGCGGCGCACTACTCTGGTGGAGTTCAAGTCCATC 540
 Db 768 ATCCACAAGGCCCTGAAGCTGAAGACGGCGGCGCACTACTCTGGTGGAGTTCAAGTCCATC 827
 QY 541 TATATGGCCCAAGAACCGCTGCAGCTGCCCGGCTACTACTACGTGGACCAACAGCTGGAC 600
 Db 828 TATATGGCCCAAGAACCGCTGCAGCTGCCCGGCTACTACTACGTGGAGTCCAACTGGAC 887
 QY 601 ATCACTCCCAACAGAGACTACACCATCTGTGAGAGTAGAGAGCGACCGAGGGCGC 660
 Db 888 ATCACTCCCAACAGAGACTACACCATCTGTGAGAGTAGAGAGCGACCGAGGGCGC 947
 QY 661 CACCACCTGTTCTCTGTA 677
 Db 948 CACCACCTGTTCTCTGTA 964

RESULT 13
 AF506025
 LOCUS
 DEFINITION Synthetic construct dimeric red fluorescent protein gene, complete cds.
 ACCESSION AF506025
 VERSION AF506025.1 GI:21464833
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM
 REFERENCE
 AUTHORS 1 (bases 1 to 681)
 Zacharias, D.A., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Campbell, R.E., and Tsien, R.Y.
 TITLE A monomeric red fluorescent protein
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7877-7882 (2002)
 MEDLINE 22056088
 PUBMED 12060735
 REFERENCE 2 (bases 1 to 681)
 AUTHORS Campbell, R.E. and Tsien, R.Y.
 TITLE Direct Submission
 JOURNAL Submitted (24-APR-2002) Pharmacology, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA
 FEATURES
 source
 1. 681
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 /note="from Discosoma sp.; Clontech vector pDsRed-N1 with human optimized codon usage"
 1. 681
 /note="dimer2; engineered variant of tetrameric red fluorescent protein from Discosoma deposited in GenBank
 Accession Number AF03369"
 /codon_start=1
 /transl_table=11
 /product="dimeric red fluorescent protein"
 /protein_id="AA054542.1"
 /db_xref="GI:21464834"
 /translation="MVASSEDVKEFMFRKVRMEGVSNGHEFIEGEGEGRPYEGTOT AKLTKGGPLPFADWILSPQFQVQSKAYVHPADIPDYKLSFPFGKRWVNFED GGVVTVQDSLSODGLIYKVRFGTNPDPGVPWOKTKWEASTERLYPRDGLKG EIHQALKDKGGHYLVEFKTIYMAKPVQLPGYVYVDTKLDITSHNEDYITVQYERS

CDS

12060735
2 (bases 1 to 1395)
Campbell, R.E. and Tsien, R.Y.
Direct Submission
Submitted (24-APR-2002) Pharmacology, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES
Location/Qualifiers
1..1395
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/notes="from Discosoma sp.; Clontech vector pDered-N1 with human optimized codon usage"
1..1395
/note="tdimer2(12); engineered variant of tetrameric red fluorescent protein from Discosoma deposited in GenBank
Accession Number AAF03369"
/codon_start=1
/transl_table=11
/product="tandem-dimer red fluorescent protein"
/protein_id="AAM54543.1"
/db_xref="GI:21464836"
/translation="MVASSEDVKEFMFKVMEGVSUNGHEFIEGEGEGRPYEGTOT
AKUVTGKGLPFWNDILSPQFGSKAYVHPADIPDYKLSFPFGFKWERNED
GGVTVTQDSSLDQGLIYKVRNKFPPDPVPMQKTMGWEASERLYPRQVLKSG
EIHQALKDKGHHLYVEFKTYNAKVPQVLPYVVDITKIDITSHNEDYITVQYERS
EGRHLFLGHGTGSGSSGTSSESDVIKEFMFKVMEGVSUNGHEFIEGEGEGRPY
YEGTQAKLVTKGGLPFWNDILSPQFGSKAYVHPADIPDYKLSFPFGFKWERNED
VMFEDG3VTVTQDSSLDQGLIYKVRNKFPPDPVPMQKTMGWEASERLYPR
DGVLKGIHQALKDKGHHLYVEFKTYNAKVPQVLPYVVDITKIDITSHNEDYITV
EQVRSRGRHLFL"

ORIGIN
Query Match 92.6%; Score 627.6; DB 12; Length 1395;
Best Local Similarity 95.7%; Pred. No. 1.9e-73;
Matches 645; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 2 TGGCCTCTCCGAGAACGTATCATCCAGGTTCATCGCCTTCAAGGTGCGATCGAGGGCA 61
Db 5 TGGCCTCTCCGAGAACGTATCATCCAGGTTCATCGCCTTCAAGGTGCGATCGAGGGCT 64
QY 62 CCCTGAACGGCCACGAGTTCGAGATCGAGGGGAGGGCGGCCCTTACGAGGGCC 121
Db 65 CCCTGAACGGCCACGAGTTCGAGATCGAGGGGAGGGCGGCCCTTACGAGGGCA 124
QY 122 ACNACACCGTGAAGTTGAAGGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 181
Db 125 CCAGACCGCCCAAGCTGAAGGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 184
QY 182 TGTCCCGCCAGTTCACGAGTTCACGAGTTCACGAGTTCACGAGTTCACGAGTTCACGAG 241
Db 185 TGTCCCGCCAGTTCACGAGTTCACGAGTTCACGAGTTCACGAGTTCACGAGTTCACGAG 244
QY 242 ACTACAGAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGTATGAATTCGAGG 301
Db 245 ACTACAGAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGTATGAATTCGAGG 304
QY 302 ACGGGGGGGTGGGACCGTGAACCCAGGACTCTCTCCCTGCGAGGAGCGGCTTCTATCTACA 361
Db 305 ACGGGGGGGTGGGACCGTGAACCCAGGACTCTCTCCCTGCGAGGAGCGGCTTCTATCTACA 364
QY 362 AGGTGAAGTTCATCGGCGTGAACCTTCCCTCCGACGGCCCGCTGATCAGAGAAGACCA 421
Db 365 AGGTGAAGTTCGCGGGGACCAACTTCCCGCCGACGGCCCGCTGATCAGAGAAGACCA 424
QY 422 TGGGCTGGGAGGCTTCCACCGAGCGCTGTACCCCGGAGCGGCTGTGAAGGGCGAGA 481
Db 425 TGGGCTGGGAGGCTTCCACCGAGCGCTGTACCCCGGAGCGGCTGTGAAGGGCGAGA 484
QY 482 TCCCAAGGGCCCTGAAGCTGAAGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 541
Db 485 TCCCAAGGGCCCTGAAGCTGAAGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 544

RESULT 14
AF506026
LOCUS
DEFINITION
AP506026 1395 bp DNA linear SYN 19-JUN-2002
Synthetic construct tandem-dimer red fluorescent protein gene,
complete cds.
AF506026 1 GI:21464835
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S.,
Zacharias, D.A., and Tsien, R.Y.
A monomeric red fluorescent protein
Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7877-7882 (2002)
22056088

ORIGIN
Query Match 92.9%; Score 629.6; DB 12; Length 681;
Best Local Similarity 95.7%; Pred. No. 1.2e-73;
Matches 647; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 2 TGGCCTCTCCGAGAACGTATCATCCAGGTTCATCGCCTTCAAGGTGCGATCGAGGGCA 61
Db 5 TGGCCTCTCCGAGAACGTATCATCCAGGTTCATCGCCTTCAAGGTGCGATCGAGGGCT 64
QY 62 CCCTGAACGGCCACGAGTTCGAGATCGAGGGGAGGGCGGCCCTTACGAGGGCC 121
Db 65 CCCTGAACGGCCACGAGTTCGAGATCGAGGGGAGGGCGGCCCTTACGAGGGCA 124
QY 122 ACNACACCGTGAAGTTGAAGGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 181
Db 125 CCAGACCGCCCAAGCTGAAGGTGACCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 184
QY 182 TGTCCCGCCAGTTCACGAGTTCACGAGTTCACGAGTTCACGAGTTCACGAGTTCACGAG 241
Db 185 TGTCCCGCCAGTTCACGAGTTCACGAGTTCACGAGTTCACGAGTTCACGAGTTCACGAG 244
QY 242 ACTACAGAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGTATGAATTCGAGG 301
Db 245 ACTACAGAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGTATGAATTCGAGG 304
QY 302 ACGGGGGGGTGGGACCGTGAACCCAGGACTCTCTCCCTGCGAGGAGCGGCTTCTATCTACA 361
Db 305 ACGGGGGGGTGGGACCGTGAACCCAGGACTCTCTCCCTGCGAGGAGCGGCTTCTATCTACA 364
QY 362 AGGTGAAGTTCATCGGCGTGAACCTTCCCTCCGACGGCCCGCTGATCAGAGAAGACCA 421
Db 365 AGGTGAAGTTCGCGGGGACCAACTTCCCGCCGACGGCCCGCTGATCAGAGAAGACCA 424
QY 422 TGGGCTGGGAGGCTTCCACCGAGCGCTGTACCCCGGAGCGGCTGTGAAGGGCGAGA 481
Db 425 TGGGCTGGGAGGCTTCCACCGAGCGCTGTACCCCGGAGCGGCTGTGAAGGGCGAGA 484
QY 482 TCCCAAGGGCCCTGAAGCTGAAGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 541
Db 485 TCCCAAGGGCCCTGAAGCTGAAGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 544
QY 542 ACATGCGCAAGAACCGGCTGAGAGTCCCGGGTACTACTAGTGGACCAAGCTGGACA 601
Db 545 ACATGCGCAAGAACCGGCTGAGAGTCCCGGGTACTACTAGTGGACCAAGCTGGACA 604
QY 602 TCACCTCCACACAGGAGGACTACACCATCGTGAGCAGTACGAGCGCACCGAGGGCGGCG 661
Db 605 TCACCTCCACACAGGAGGACTACACCATCGTGGAACAGTACGAGCGCTCCGAGGGCGGCG 664
QY 662 ACCACCTGTTCTGTGTA 677
Db 665 ACCACCTGTTCTGTGTA 680

```
QY 542 ACATGGCCAAAGACCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACA 601
Db 545 ACATGGCCAAAGACCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACA 604
QY 602 TCACCTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGCGCGCC 661
Db 605 TCACCTCCCAACAGAGGACTACACCATCGTGGAGCAGTACGAGCGCTCCGAGCGCGCC 664
QY 662 ACCACCTGTCCTG 675
Db 665 ACCACCTGTCCTG 678

RESULT 15
AX233584
LOCUS AX233584 681 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 8 from Patent WO0162919.
ACCESSION AX233584
VERSION AX233584.1 GI:15593307
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
          artificial sequences.
REFERENCE
1 Nelson,D., Zamaira,E. and Tsien,R.
  Modified fluorescent proteins
  TITLE Patent: WO 0162919-A 8 30-AUG-2001;
  JOURNAL Aurora Biosciences Corporation (US)
FEATURES
source
1..681
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="Mutant Anthozoan red fluorescent protein"
1..681
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC69734.1"
/db_xref="GI:15593308"
/db_xref="RENTREMBL:CAC69734"
/translation="MVRSSKNVKEFMRPKVRMEGTUNGHEFIEGEGSRPYEGHNT
VKLVTKGGPLFPADILSPQYIGSKVYVHPADIPDYKKLSFPEGFRWRWMPED
GGVVTVDQSSLDQGFYKVFYGVNFPDGPVQKMGWEASTERYLPDRGVKLG
EIKKALKDKDGHYLVBFKSIYMAKRPVLPGYYVDSKLIDITSHNEDYTIIVEQVET
EGRHHLFL"

CDS
Query Match 90.6%; Score 614.2; DB 6; Length 681;
Best Local Similarity 95.1%; Pred. No. 1.3e-71;
Matches 634; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 12 CGAGAACGTATACCGAGTTCATCGCTTCAGGTGCGCATGGAGGGACCGTGAACGG 71
Db 15 CAAGACGTGATCAAGAGTTCATGAGTTCAGGTGCGCATGGAGGGACCGTGAACGG 74
QY 72 CCACAGTTCAGATCGAGGGCGAGGGCGGCCGCCCTACGAGGGCCACAACCCGT 131
Db 75 CCACAGTTCAGATCGAGGGCGAGGGCGGCCGCCCTACGAGGGCCACAACCCGT 134
QY 132 GAAGTTGAAGTGACCAAGGGCGGCCGCCCTTCGCTCGGAGCATCTGTCCCCCA 191
Db 135 GAAGCTTAAGTGACCAAGGGCGGCCGCCCTTCGCTCGGAGCATCTGTAGCCCCCA 194
QY 192 GTTCCAGTACGGCTCCAAAGTGTAGTGAAGCAGCCCGCCGACATCCCGACTACAAGAA 251
Db 195 GTTCCAGTACGGCAGCAAGGTGTACGTGAAGCAGCCCGCCGACATCCCGACTACAAGAA 254
QY 252 GCTCTCTCCCGAGGGCTTCAGTGGAGCGGTGATGAATTCGAGGACGGCGGT 311
Db 255 GCTGAGCTTCCCCGAGGGCTTCAGTGGAGGAGGTGATGAATTCGAGGACGGCGGT 314
QY 312 GGCACCGTGACCCAGGACTCTCTCCCTGCAGGAGCGGTCTTCTCATCAAGGTGAAGTT 371
```

```
Db 315 GGTGACCGTGACCCAGGACAGAGCGCTGCAGACGGCTGCTTTCATCTACAGGTGAAGTT 374
QY 372 CATCGCGGTGAACATTCCCTCCGACGGCCCGTGTATGAGAGAGACCAATGGGCTGGGA 431
Db 375 CATCGCGGTGAACATTCCCTCCGACGGCCCGTGTATGAGAGAGACCAATGGGCTGGGA 434
QY 432 GGCCTCCACCGAGCGCTGTACCCCGGAGCGGTGCTGAAGGGCGAGATCCACAAGGC 491
Db 435 GGCCTCCACCGAGCGCTGTACCCCGGAGCGGTGCTGAAGGGCGAGATCCACAAGGC 494
QY 492 CCTGAAGCTGAAGGACGGCGCCACTACTCTGTGGAGTTCAGTCCCATCTACATGGCAA 551
Db 495 CCTGAAGCTGAAGGACGGCGCCACTACTCTGTGGAGTTCAGTCCCATCTACATGGCAA 554
QY 552 GAAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACATCACTCCCA 611
Db 555 GAAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACTCCAAAGCTGGACATCACTCCCA 614
QY 612 CAACGAGGACTACACCATCGTGGAGCAGTACGAGCGACCGAGGGCGGCCACCTGTT 671
Db 615 CAACGAGGACTACACCATCGTGGAGCAGTACGAGGAGGACCGAGGGCGGCCACCTGTT 674
QY 672 CCTGTAA 678
Db 675 CCTGTGA 681

Search completed: August 3, 2004, 14:04:16
Job time : 4537.5 secs
```

Sequence 12, Appl
Sequence 17, Appl
Sequence 13, Appl
Sequence 18, Appl
Sequence 12, Appl
Sequence 17, Appl
Sequence 9, Appl
Sequence 28, Appl
Sequence 33, Appl
Sequence 28, Appl
Sequence 33, Appl
Sequence 30, Appl
Sequence 30, Appl
Sequence 33, Appl
Sequence 3, Appl
Sequence 5, Appl
Sequence 31, Appl
Sequence 21, Appl

Sequence 12, Appl
Sequence 17, Appl
Sequence 13, Appl
Sequence 18, Appl
Sequence 12, Appl
Sequence 17, Appl
Sequence 9, Appl
Sequence 28, Appl
Sequence 33, Appl
Sequence 28, Appl
Sequence 33, Appl
Sequence 30, Appl
Sequence 30, Appl
Sequence 33, Appl
Sequence 3, Appl
Sequence 5, Appl
Sequence 31, Appl
Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-459-956-6
; Sequence 6, Application US/09459956
; Patent No. 6342372
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, Iii, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-09-459-956-6

Query Match 64.6%; Score 438; DB 4; Length 678;
Best Local Similarity 77.9%; Pred. No. 3.8e-76;
Matches 528; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 1 ATGGCTCTCCGAGACGTCATCACCGAGTTTCATCGCTTCAAGGTGCGCATGAGGC 60
DB 1 ATGAGGCTTCCCAAGATGTTATCAAGAGTTTCATGAGTTTAAAGTTTCGATGAAGGA 60
QY 61 ACCGTGAACGCCACGAGTTTCAGATCGAGGCGGAGGCGGAGGCGCCCTTACGAGGC 120
DB 61 ACCTGCAATGGCGACGAGTTTGAATAGAGCGGAGGAGGCGGAGGCGCCATACGAGGC 120
QY 121 CACACACCGTGAAGTTGAGGTGACCCAGGCGGCGCCCTTCCCTTCCCTGGGACATC 180
DB 121 CACATACCGTGAAGTTTAAAGTAAACCAAGGCGGAGCTTTGCCATTTCCTTGGGATAT 180
QY 181 CTGTCCCGCCAGTTCCAGTACGCTTCAAGGTGACGTGAAGCACCCCGCGCATCCCC 240
DB 181 TTGTCCACCAATTTAGTATGGAAGCAGAGTATATGTCAAGCACCTTCCGACATACCA 240
QY 241 GACTACAAAGAGCTGTCTTCCCGAGGCTTCAAGTGGGAGCGGCTGATGAATTCGAG 300
DB 241 GACTATAAAGAGCTGTCTTCCCTCAAGAGGATTTAAATGGGAAGGCTCATGAATTCGAA 300

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 11:04:09 ; Search time 68.5343 Seconds
(without alignments)
5490.050 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 678
Sequence: 1 atggctctccgagacgt.....gccaccacgttctctgttaa 678

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
6: /cgn2_6/prodata/2/ina/5B_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	438	64.6	678	4	US-09-459-956-6
2	257.4	38.0	699	4	US-09-459-956-5
3	251.8	37.1	720	4	US-09-839-650-1
4	216.6	31.9	801	4	US-09-459-956-7
5	144.8	21.4	690	4	US-09-459-956-2
6	119	17.6	1104	3	US-09-277-716-30
7	119	17.6	1104	4	US-09-609-161B-30
8	119	17.6	1279	3	US-09-277-716-31
9	119	17.6	1279	4	US-09-609-161B-31
10	110.4	16.3	696	4	US-09-459-956-4
11	100.2	14.8	717	4	US-09-214-909-1
12	99.2	14.6	696	4	US-09-459-956-3
13	99	14.6	717	4	US-09-486-241-31
14	98.4	14.5	1079	4	US-09-609-161B-15
15	98.4	14.5	1079	4	US-09-626-581B-64
16	98.4	14.5	1079	4	US-09-415-765B-64
17	98.4	14.5	1079	4	US-09-626-580C-64
18	98.4	14.5	1085	4	US-09-277-716-15
19	95.2	14.0	1021	4	US-09-839-650-2
20	94.4	13.9	782	1	US-08-532-390-40
21	94.4	13.9	782	3	US-08-717-294-40
22	93	13.7	720	3	US-09-172-063-14
23	93	13.7	720	4	US-09-316-919-15
24	93	13.7	720	4	US-09-602-641-14
25	92.8	13.7	7353	4	US-08-786-531B-1
26	91.4	13.5	717	4	US-09-513-783A-43
27	91.4	13.5	720	3	US-09-094-359-9

QY 301 GACGGCGCGTGGCGACCGTGAACCCAGGACTCCTCCCTGCGAGGACGGCTCTTCATCTAC 360
DB 301 GACGGTGGCGTGTCTGATGATACCCAGGATTCAGTTGCGAGATGGCTCTTCATCTAC 360
QY 361 AAGGTGAAGTTCATCGCGTGAATCTCCCTCCGACGGCCCCCGTGAATGCAAGAAAGAC 420
DB 361 AAGGTGAAGTTCATCGCGTGAATCTCCCTCCGACGGCCCCCGTGAATGCAAGAAAGAC 420
QY 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGGCGTGTGAAGGGCGAG 480
DB 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGGCGTGTGAAGGGCGAG 480
QY 481 ATCCACAGGCGCTGAAGCTGAAGGACGGCGGCGCACTACCTGTGGAGTTCAAGTCCATC 540
DB 481 ATTCAATAGGCTCTGAAGCTGAAGGACGGTGGTCACTTACCTAGTTGAATTCAAAGTAAT 540
QY 541 TACATGGCAAGAACCGCGTGCAGCTCCCGGCTACTACTACGTGGACCAACAGCTGGAC 600
DB 541 TACATGGCAAGAACCGCGTGCAGCTCCCGGCTACTACTACTACGTGGACCAACAGCTGGAT 600
QY 601 ATACCTCCCAACAGGAGGACTACACCATCGTGAGCAGTACGAGCGCACCGAGGGCGCG 660
DB 601 ATACAGGACCAACAGGAGGACTACACCATCGTGAGCAGTATACATCGTTGAGCAGTATGAAGAACCGAGGACGC 660
QY 661 CACCCTGTTCCTGTAA 678
DB 661 CACCCTGTTCCTTTAA 678

RESULT 2

US-09-459-956-5
; Sequence 5, Application US/0945956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR FILING DATE: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Discosoma striata
US-09-459-956-5

Query Match 38.0%; Score 257.4; DB 4; Length 699;
Best Local Similarity 62.4%; Pred. No. 2.1e-41;
Matches 421; Conservative 0; Mismatches 251; Indels 3; Gaps 1;

QY 1 ATGCGCTCTCCGAGAACGTCATCCGAGTTCATCGGTTCAAGTGGCATGGAGGC 60
DB 1 ATGAGTTGTTCCAAAGAGTGTGATCAAGGAAGAAATGTTGATCATCTTCACTGGAAGA 60
QY 61 ACCTGACCGCCACAGTTCGAGATCGAGGCGGCGGCGCCCTACGAGGC 120
DB 61 ACCTGACCGCCACAGTTCGAGATCGAGGCGGCGGCGCCCTACGAGGC 120
QY 121 CACAACACCGTGAAGTTGAAGTGAACCAAGGGGGCGCCCTCGCCCTTCGCTGGGACATC 180
DB 121 ACCAATACCGTCAAGTTCGAGTTACCAAGGGTGGACCTCTGCCATTTGTTGGCATATT 180
QY 181 CTGTCCCCCAGTTCAGTACGCTCCAGGTGTAGTGAAGCACCCCGCCGACATCCCC 240

DB 181 TTGTGCCCAATTTTCAGTATGGAACAAGSCATTTTGTCCACCACCCTTGACACATCAT 240
QY 241 GACTACAGAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGAGCGGTGATGAATCTTGAG 300
DB 241 GATTATCTAAGAGCTGTCTCTCCCGAGGGATATACATGGGAACGGTCCATGCACTTTGAA 300
QY 301 GACGGCGCGTGGCGACCGGTGACCCAGGACTCCTCCCTGCGAGGACGGTGTCTCATCTAC 360
DB 301 GACGGTGGCTTGTGTGTATCAACATGATATCAGTTTGACAGCAACTGTTTCTACTAC 360
QY 361 AAGGTGAAGTTCATCGCGTGAATCTCCCTCCGACGGCCCCCGTGAATGCAAGAAAGAC 420
DB 361 GACATCAAGTTCATCGCGTGAATCTCCCTCCGACGGCCCCCGTGAATGCAAGAAAGAC 420
QY 421 ATGGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGGCGTGTGAAGGGCGAG 480
DB 421 ACTGGCTGGAGAACCGAGCACTGAGCGTTTGTATCTCTGTATGTTGTGTATAGGAGAC 480
QY 481 ATCCAAAGGCGCTGAAGCTGAAGGACGGCGGCGCACTACCTGTGGAGTTCAAGTCCATC 540
DB 481 ATCCATCATGCTCTGACAGTTGAAGGAGTGTCTATACGCATGTGACATTAATAACTGTT 540
QY 541 TACATGGCAAGAACCGCGTGCAGCTCCCGGCTACTACTACGTGGACCAACAGCTGG 597
DB 541 TACAGGCGCAAGAAAGGGCGCTTGAAGATGCCAGGGTATCACTATGTTGACACCAACTG 600
QY 598 GACATCACCTCCCAACAGGAGGACTACACCATCGTGAGCAGTACGAGCGCACCGAGGC 657
DB 601 GTTATATGGAACAACGACAAAGATTCATGAAGTTGAGGAGCATGAATCCGCCGTTGCA 660
QY 658 GCCACACCTGTTTC 672
DB 661 CGCCACCATCGGTTTC 675

RESULT 3

US-09-839-650-1
; Sequence 1, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; TITLE OF INVENTION: Fluorescent Protein
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Humanized R. mulleri polynucleotide
; NAME/KEY: misc feature
; LOCATION: (1)-(720)
; OTHER INFORMATION: Humanized DNA sequence
US-09-839-650-1

Query Match 37.1%; Score 251.8; DB 4; Length 720;
Best Local Similarity 62.4%; Pred. No. 2.6e-40;
Matches 394; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

QY 28 GAGTTCATCGCTTCAGGTTCCGATCGAGGCGACCGTGAACGGCCACGAGTTCCGATC 87
DB 40 GAGGTGATGAGTCAAGGTGAACCTGGAGGCGATCGTGAACCAACCCAGGTGTTCCCATG 99
QY 88 GAGGGCGAGGCGAGGCGCGCCCTACGAGGGCCACCAACACCGTGAAGTTGAAGGTGACC 147
DB 100 GAGGGCTCGGCAAGGCGCAACATCTGTTCCGCAACACGAGCTGTTGCGAGATCCCGGTGACC 159
QY 148 AAGGGCGGCGCCCTTGGCCCTTGGCATCTGTTCCCGCCAGTTCAGTACGAGTCC 207

Db 160 RAGGGGCCCCCTTGCCTTCGATCGTGGCCCGCTTCCAGTACGGCAAC 219
Qy 208 AAGGTGTAGTGAAGACCCCGGACATCCCGGACTCAAGAACTGTCTTCCCGAG 267
Db 220 CGCACCTTACCAGTACCCCAACGACATCAGCGACTTTCATCCAGAGCTTCCCGCC 279
Qy 268 GGTTCCTCAAGTGGAGCGGTGATGAATTCGAGGACGGGGGCTGGCGACCGTGAACCCAG 327
Db 280 GGTTCATCTACGAGCGCACCTTGGCTACGAGGACGGGGCTGGTGGAGATCCGAGC 339
Qy 328 GACTCTCTCTGAGAGCGGTGCTTTCATCTCAAGGTGAAGTTCATCGCGGTGAACCTTC 387
Db 340 GACATCAACCTGATCGAGGACAAAGTTCGTGTACCGCGTGGAGTCAAGAGGCGACAACTTC 399
Qy 388 CCTCCGAGCGCCCGGTGATGAGAGGAGAGACCATGGGTGGAGGCGCTCCACCGAGCGC 447
Db 400 CCGAGCAGCGCCCGGTGATGAGAGGAGACCATCTGGGATCGAGCCAGCTTCGAGGCC 459
Qy 448 CTGTACCCCGGAGCGGTGCTGTAAGGCGGAGATCCACAGGCGCTGAAGCTGAAGGAC 507
Db 460 ATGTATCATGAACAAACCGGTGCTGGTGGCGGAGGTGATCTGGTGTACAGCTGAACAGC 519
Qy 508 GGGGCGCACTACTGCTGAGTTCAGTCCATCTACATGCGGCAAGAGCCCGTGCAGCTG 567
Db 520 GGCAAGTACTACAGTGCACATGAAGACCTGTATGAAGAGCAAGGCGGTGTGAAGGAG 579
Qy 568 CCGGGTACTACTACGTGAGACCAAGCTGGACATCACCTCCCAACAGGAGACTACACC 627
Db 580 TTCCCTCTACACCTTCTACGACACCGCTGGAGAGACCTACGTGAGGAGCGCGGC 639
Qy 628 ATCGTGGAGCTACGAGCGCACCGAGGCC 658
Db 640 TTCGTGGAGCAGCAGAGACCGCCATCGCCC 670

RESULT 4
US-09-459-956-7
; Sequence 7, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Clavularia sp
US-09-459-956-7

Query Match 31.9%; Score 216.6; DB 4; Length 801;
Best Local Similarity 58.2%; Pred. No. 1.5e-33;
Matches 381; Conservative 0; Mismatches 274; Indels 0; Gaps 0;
Qy 8 CTTCCGAGAACGTCATCCAGGTTTCATCGGTTCAAGTGGCGCATGGAGGCGACCGTGA 67
Db 122 CCACCATGGGTGTGATTAACACGACATGAAGATTAAGCTGAAGTGAAGGAATGTAA 181
Qy 68 ACGCCACAGTTCGAGATCGAGGCGGAGGCGGCGCCCTACGAGGCGCACCA 127
Db 182 ACGGGATGCTTTTGTGATGAAGGAGAGGAGAAAGCCCTTACGATGGACACACA 241

Qy 128 CCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTGCCTTCCGCTGGGACATCCTGTCCC 187
Db 242 CTTTAAACCTTGAAGTGAAGAGGTGGCGCTCTGCCTTTTCTTACGATATCTTGTCAA 301
Qy 188 CCGAGTTCCAGTACGGCTCAAGGTGTACGTGAAGCAACCCCGGCGACATCCCGAGTACA 247
Db 302 ACGGTTCCAGTACGGAAACAGAGCATTTGACAAATACCCAGACGATATAGAGACTATT 361
Qy 248 AGAAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCTGATGAACCTTCGAGAGCGCG 307
Db 362 TCAAGCAGTGTCTTCCCGAGGGATATCTCGGAGAAACCATGACTTTTGAAGACAAAG 421
Qy 308 GCGTGGCGACCGTGAACCCAGGACTCTCTCGAGGACGCTGCTTCATCTCAAGGTGA 367
Db 422 GCATTGTCAAAAGTGAAAGTGACATAAGCATGGAGGAAGACTCCTTTTATCTATGAATTC 481
Qy 368 AGTTCATCGCGGTGAACCTTCCCTCCGAGCGCCCGGTGATGACAGAAAGACCATGGCT 427
Db 482 GTTTTGTAGGATGAACCTTCTCCCAATGCTCGGTTATGCGAAGAAACCTTTGAAGT 541
Qy 428 GGGAGGCTCTCACCGAGCGCTGTACCCCGGACGCGCTGTGAAGGGCGAGATCCACA 487
Db 542 GGGAAACCATCCACTGAGATTATGTACGTGCTGATGGAGTGTGTCGAGATATTAGCC 601
Qy 488 AGGCCCTGAAGCTGAAGAGCGGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGG 547
Db 602 ATTCTCTGTTGCTGGAGGAGGTGGCCATTACCGATGTGACTTCAAAAGTATTTACAAAG 661
Qy 548 CCAAGAACCGCTGACGCTGCCCGCTACTACTACGTGGACACCAAGCTGGACATCACCT 607
Db 662 CAAAAAAGTTGTCAAAATTCGAGACTATCACTTTGGACCATCGCATTTGAGATCTGA 721
Qy 608 CCCCAACGAGGACTACACCATCGTGGAGAGTACGAGCGGACCGAGGGCGGCCA 662
Db 722 ACCATGACAGGATTACAAAGTAACGCTGTATGAGAAATGCAAGTTGCTCGCTA 776

RESULT 5
US-09-459-956-2
; Sequence 2, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Anemonia majano
US-09-459-956-2

Query Match 21.4%; Score 144.8; DB 4; Length 690;
Best Local Similarity 52.4%; Pred. No. 9.8e-20;
Matches 345; Conservative 0; Mismatches 307; Indels 6; Gaps 1;
Qy 1 ATGGCTCTCTCCGAGAACGTCATCCAGGTTTCATCGGTTCAAGGTGGCGCATCGAGGC 60
Db 1 ATGGCTCTTCAACACAGTTTATCGAGATGACATGAATGACCTACCATATGATGGTGC 60
Qy 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGGAGGCGCGCCCTTACGAGGC 120

Db 61 TGTGTAATGGGCAATTAACGTTCAAGGTGAAGGCAACGGGAAGCATTACGAAGG 120
Qy 121 CACACACCGTGAAGTTGAAGTGCACCAAGGC-----GGCCCCCTGGCTCGCTGG 174
Db 121 ACGCAGACTTCGACTTTTAAAGTCAACCATGCGGAGGCGGCTTGCATTCTCTTT 180
Qy 175 GACATCTGTCCCGCCAGTTCCAGTACGGTCCCAAGGTGACGTGAAGCACCCCGCCGAC 234
Db 181 GACATATCTACAGTGTTCAAATATGAATCGATGCTTTACTCGTATCTACCACT 240
Qy 235 ATCCCGCACTACAGAGTGTCTTCCCGAGGGTTCAGTGGAGGCGGTGATGAC 294
Db 241 ATGCCCGACTATTTCACCAAGCATTTCTGACGGAATGTCATGAAAGGACTTTTACC 300
Qy 295 TTGAGGACGGCGGTGGGCGACCGTGACCCAGGACTCTCTCCCTGAGGACGGCTCTTC 354
Db 301 TATGAAGATGGAGGTTGCTACAGCCAGTTGGGAATAAGCCTTAAGGCACTGCTTT 360
Qy 355 ATCTAAGGTGAAGTTCACTCGCGTGAATTCCTCCCTCCAGCGGCCCGCGTGAAGA 414
Db 361 GAGCACAATCCAGTTTTCATGAGTGAATTTCTCTGCTGATGACCTGTGATGGCGA 420
Qy 415 AAGACCATGCGTGGGAGGCTCCACCGAGCGCTGTACCCCGGACGCGGTGCTGAAG 474
Db 421 AAGACAATGTTGGGACCATCTTTGAGAAATGACTGTCTGCGATGGAATATTGAAG 480
Qy 475 GCGGAGATCCACAGGCCCTTGAAGTGAAGAGCGGCGGCACTACCTGTGAGTTCAAG 534
Db 481 GGTGATGTCACCGGTTCTCTCATGCTGCAAGAGGTGGCAATTACAGATGCCAATCCAC 540
Qy 535 TCCATCTACATGCGCAGAGCGGTGACGTGCGCGGTCTACTACTGTGACGACCAAG 594
Db 541 ACTTCTTACAGACAAACCGGTGACGATGCCACCAACCATGTGTGGAAACATGCG 600
Qy 595 CTGACATCACTCCCAACAGGAGGACTACACCATGCTGGAGCAGTACGAGGCGCACCG 652
Db 601 ATTGCGAGGACCGACCTTGACAAAGGTGGCAACAGTGTTCAGCTGACGAGGACGCTG 658

RESULT 6

US-09-277-716-30
; Sequence 30, Application US/09277716A
; Patent No. 6232107

GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce

; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26

; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01

; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15

; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 30
; LENGTH: 1104

; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi

; FEATURE:
; NAME/KEY: CDS

; LOCATION: (34)..(747)
; FEATURE:

; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-277-716-30

Query Match

Best Local Similarity 17.6%; Score 119; DB 3; Length 1104;

Matches 308; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

Qy 23 TCACCGAGTTTCATCGGCTTCAAGGTGGCGCATGGAGGCGACCGTGAACGCGCCAGGTTCC 82
Db 65 TGAAGAGATTATGTCGCGCAAAAGCTAGCGTTGAAGGAATCGTGAACAAATCACGTTTTTT 124
Qy 83 AGATCGAGGCGGAGGCGGAGGCGCCCTTACGAGGCGCACACACCGTGAAGTTGAAGG 142
Db 125 CCATCGGAAGGATTTGGAAGGCAATGTAATATTTTGAAGAACCAATGATGCAAAATCCGG 184
Qy 143 TGACCAAGGCGGCGCCCTTGCCTTCCCTGGGACATCTGTCTCCCGCAGTTCAGTACG 202
Db 185 TTACAGAGGAGGTCCTTGCCTTCCGTTTCGATATTTGTTCCATAGCTTTCCAAATACG 244
Qy 203 GCTCCAGGTTGACGTGAAGCACCCCGCGGACATCCCGGACTCAAGAAAGTGTCTCTTC 262
Db 245 GGAATCGCACTTTCACGAAATACCCAGACGACATTGGCGACTACTTTGTTCAATCATTC 304
Qy 263 CCGAGGCGTTCAAGTGGGAGCGCGTGATGAACCTTCGAGGAGCGGCGGTGGCGACCGTGA 322
Db 305 CGGCTGANTTTTCTACGAAGAATCTACGCTTTGAAGATGGCGCCATTTGTTGACATTC 364
Qy 323 CCCAGGACTCTCCCTGCGAGCGGCTGCTTCACTTACAAGGTGAAGTTCAATGGCGTGA 382
Db 365 GTTCAGATATTAAGTTTGAAGATGATAAGTTCCACTACAAAGTGGAGTAGAGGCAACG 424
Qy 383 ACTTCCCTCCGAGCGCGCGTGATGACAGAGACCATGGCTGGGAGGCGCTCCACCG 442
Db 425 GTTTCCTTAGTAAAGCGCGGTGATGCAAAAGGCGCATCTCCGCGATGGAGCCATCGTTG 484
Qy 443 AGCGCTGTATACCCCGGAGCGGCTGTGAAGGCGGAGATCCCAAGGCGCTGAAAGTGA 502
Db 485 AGTGGTCTACATGAACAGCGGCGTCTGGTGGGGAAGTAGATCTGTTTACAACTCG 544
Qy 503 AGGAGCGGCGGCGCTACTGTTGGAGTTCAAGTCCATCTACATGCGGCGCAAGAGCCGTCG 562
Db 545 AGTCAGGGAACATTAATCTGTCGCGACATGAAAGGTTTACAGATCCAAAGGTGGAGTGA 604
Qy 563 AGCTGCGCGGCTACTACTAGTGGACACCAAGCTGGACATCACCTCCCAACAGGACT 622
Db 605 AAGAAATCCCGAATATCATCTTATCCATCATCTGCTGGAGAAACCTACGTGGAAGAG 664

RESULT 7

US-09-609-161B-30

; Sequence 30, Application US/09609161B

; Patent No. 6436682

GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce

; APPLICANT: Szent-Gyorgyi, Christopher

; APPLICANT: PROLUME, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B

; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26

; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01

; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15

; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 30
; LENGTH: 1104

; TYPE: DNA

ORGANISM: Ptilosarcus gurneyi
FEATURE:
NAME/KEY: CDS
LOCATION: (34)...(747)
OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-609-161B-30

Query Match 17.6%; Score 119; DB 4; Length 1104;
Best Local Similarity 49.4%; Pred. No. 9.2e-15;
Matches 308; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

Qy	23	TCACGAGTTTCATGCGCTTCACAGTTCAGGTGCGCATGAGGCGCACCGTGAACGCGCACAGATTTCG	82
Db	65	TGAAGAGATTATGTCGCGCAAGCTAGCGTTGAAGGAATCGTGAACAAATCACGTTTTTT	124
Qy	83	AGATCGAGGCGGAGGCGGCGCGCCCTACGAGGCGCACACACCGTGAAGTTGAAGG	142
Db	125	CCATGGAAGATTGGAAGGCAATGATATTTTGGAAACCAATTGATGCAATCCGG	184
Qy	143	TGACCAAGGCGGCGCCCTGCGCTGCGCTGGGAATCTGTCGCCCCAGTTCCAGTACG	202
Db	185	TTACAAAGGAGGTCGTTGCCATTCGCTTTCGATATTTTCCATAGCTTTCCAAATACG	244
Qy	203	GCTCCAGGTGTACGTGAAGCACCAGCGGACATCCCGACATCCCGACATCAAGAGCTGCTCTCC	262
Db	245	GGATCGCACTTTCAGGAATACCCAGACGACATTCGGGACTCTTTGTTCAATCATTCC	304
Qy	263	CCGAGGCTTCAAGTGGAGCGCGTGAATGAATTCGAGGACGCGCGGCTGGCGACCGTGA	322
Db	305	CGGCTGGAATTTTCTACGAAAGAAATCTACGCTTTCAAGATGGCGCAATGTTGACATTC	364
Qy	323	CCGAGACTCTCCCTGAGGAGCGCTGCTTCTATCTACAGGTGAATTCATCGCGGTA	382
Db	365	GTTCAATATAAGTTTGAAGATGATAAGTTCCACTTACAAAGTGGATATAGGCAACG	424
Qy	383	ACTTCCCTCCGAGCGCCCTGATGACAGAAAGACCATGGGCTGGAGGCGCTCCACCG	442
Db	425	GTTTCCCTAGTAAAGGACCGCTGATGACAAAGCCATCTCGCATGGAGCCATCGTTG	484
Qy	443	AGGCGCTGTACCCCGCGAGCGGCTGCTGAAGGCGAGATCCACAGGCGCTTGAAGCTGA	502
Db	485	AGTGCTGTACATGAACAGCGGCTTCTGCTGGCGAAGTAGATCTCGTTTCAAACTCG	544
Qy	503	AGGAGCGGCGCACTACCTGCTGGAGTTCAAGTCCATCTACATGGCCAGAGAGCGCGTC	562
Db	545	AGTCAGGAACTATTAATCTGTCGCCACATGAAACGTTTACAGATCCAAAGTGGAGTGA	604
Qy	563	AGTGCGCGGCTACTACTAGTGGACACCAAGCTGGACATACCTCCCAACAGGAGACT	622
Db	605	AAGAAATCCCGGAATATCACTTTATCCATCATCTGCTGGAGAAACCTACGTGGAAGAG	664
Qy	623	ACACCATCGTGGAGCTAGGAG	645
Db	665	GAAGCTTCGTGGAAACACGAG	687

RESULT 8

US-09-277-716-31
Sequence 31, Application US/09277716A
Patent No. 6232107
GENERAL INFORMATION:
APPLICANT: Bryan, Bruce
APPLICANT: Szent-Gyorgyi, Christopher
APPLICANT: PROMUE, LTD.
TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
CURRENT APPLICATION NUMBER: US/09/277,716A
CURRENT FILING DATE: 1999-03-26
EARLIER APPLICATION NUMBER: 60/102,939
EARLIER FILING DATE: 1998-10-01
EARLIER APPLICATION NUMBER: 60/089,367
EARLIER FILING DATE: 1998-06-15
EARLIER APPLICATION NUMBER: 60/079,624
EARLIER FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 31
LENGTH: 1279
TYPE: DNA
ORGANISM: Ptilosarcus gurneyi
FEATURE:
NAME/KEY: CDS
LOCATION: (7)...(720)
OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert B)
US-09-277-716-31

Qy	23	TCACGAGTTTCATGCGCTTCAAGTTCGCGCATGAGGCGCACCGTGAACGCGCACAGATTTCG	82
Db	38	TGAAGAGATTATGTCGCGCAAGCTAGCGTTGAAGGAATCGTGAACAAATCACGTTTTTT	97
Qy	83	AGATCGAGGCGGAGGCGGCGCGCCCTACGAGGCGCACACACCGTGAAGTTGAAGG	142
Db	98	CCATGGAAGATTTCGGAAGGCAATGATATTTTGGAAACCAATTGATGCAATCCGG	157
Qy	143	TGACCAAGGCGGCGCCCTGCGCTTGGCGCATCTCTGCTCCCGCTGCTCCCGCTGCTCCAGTACG	202
Db	158	TTCAAAGGAGGTCGTTGCCATTCGCTTTCGACATGTTTCCATAGCTTCCAAATACG	217
Qy	203	GCTCCAGGTGTACGTGAAGCACCAGCGGACATCCCGGACTGCTCCGACTCAAGAGCTGCTCTCC	262
Db	218	GGATCGCACTTTCACGAAATACCCAGACGACATTCGGGACTACTTTTGTTCATCATTC	277
Qy	263	CCGAGGCTTCAAGTGGAGCGGCTGATGAATTCGAGGAGCGGCGCTGCGGACCGTGA	322
Db	278	CGGCTGATTTTCTACGAAAGAAATCTACGCTTGAAGATGGCGCCATTTGACATTC	337
Qy	323	CCGAGGACTCTCCCTCGAGGACCGCTGCTTCTATCTACAAGGTGAAGTTTCAATCGCGCTGA	382
Db	338	GTTCAAGATATAAGTTTAGAAGATGATAAGTTCCACTACAAGTGGACTATAGAGCAACG	397
Qy	383	ACTTCCCTCGAGCGGCGCTGATGAGAAAGACCATGCGGTGGAGGCGCTCCACCG	442
Db	398	GTTTCCCTAGTAAAGGACCGCTGATGCAAAAGCCATCTCGGATGAGGCGCTGCTGTTG	457
Qy	443	AGCGCTGTATCCCGCGGAGCGGCTGCTGAAGGCGGAGATCCCAAGGCGCTGAAGCTGA	502
Db	458	AGTGCTTCTATGAACAGCGGCTTCTGTTGGCGGAGTAGATCTCGTTTACAACCTCG	517
Qy	503	AGGAGCGGCGGCTACTACTGTTGAGTTCAAGTCCATCTACATGCGCCAAAGCCCGTGC	562
Db	518	AGTCAGGAACTATTACTCGTCCACATGAAACGTTTTCAGATCCAAAGGTGGAGTGA	577
Qy	563	AGCTGCGGCTACTACTAGTGGACACCAAGCTGGACATCACCTCCCAACAGGAGACT	622
Db	578	AAGAATTCGCGGAATATCACTTTATCCATCATCTCTGGAGAAACCTACGTGGAAGAG	637
Qy	623	ACACCATCGTGGAGCTAGGAG	645
Db	638	GAAGCTTCGTGAACACACGAG	660

RESULT 9

US-09-609-161B-31
Sequence 31, Application US/09609161B
Patent No. 6435682
GENERAL INFORMATION:
APPLICANT: Bryan, Bruce
APPLICANT: Szent-Gyorgyi, Christopher
APPLICANT: PROMUE, LTD.
TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIC
TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS

```
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 31
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(720)
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert B)
US-09-609-161B-31

Query Match      17.6%; Score 119; DB 4; Length 1279;
Best Local Similarity 49.4%; Pred. No. 9.3e-15; Indels 0; Gaps 0;
Matches 308; Conservative 0; Mismatches 315;

Qy 23 TCACCGAGTTTCATCGGCTTCAAGTGGCGATGAGGCGCACCGTGAACGCGCCACGAGTTCCG 82
Db 38 TGAAGAGATATGTCGGCAAAAGTAGCGTTGAAGGAATCGTGAACAATCACGTTTTTT 97

Qy 83 AGATCGAGGCGGAGGCGGCGCGCCCTCAGAGGCGCACACACCGTGAAGTTGAAG 142
Db 98 CCATCGGAAGATTGGAAGAGGCAATGATATTATTGGAACCAATGATGCAATCCGG 157

Qy 143 TGACCAAGGCGGCGCGCCCTCCCTTCGCTGGGACATCTGTCGCCCGGAGTTCCAGTAG 202
Db 158 TTACAAAGGAGGTCGTTGCCATTCGTTTCGACATTGTTTCCATAGCTTTCCATAG 217

Qy 203 GCTCAAGGTTAGTGAAGCACCCCGCGACATCCCGGACTACAGAGAGTGTCTCTCC 262
Db 218 GGAATCGCACTTTCACGAAATACCCAGACGACATTCGGACTACTTTGTTCAATCAT 277

Qy 263 CCGAGGCGTTCAAGTGGGCGGCGTGAATGAATTCGAGGAGCGCGCGTGGCGACCGTGA 322
Db 278 CGGCTGGATTTTCTAGCAAGAAATCTAGCTTTGAAGTGGCGCATTTGTGACATTC 337

Qy 323 CCAGAGGACTCCTCCCTCGAGAGCGGTGCTTCAATCAAGTGAAGTCAATCGGGCTGA 382
Db 338 GTTCAGATATAAGTTTGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGGCAAG 397

Qy 383 ACTTCCCTCCGAGCGCGCGTGCATCAGAGAGACCATGGCTGGAGGCTCCACCG 442
Db 398 GTTTCCTTAGTACGAGCGCGTGCATGCAAAAGCCATCTCGGATGGAGCCATCTTTTG 457

Qy 443 AGCGCTGTATCCCGCGAGCGGCTGCTGAAGGCGGAGATCCCAAGGCGCTGAAGCTGA 502
Db 458 AGGTGGTCTCATGAACAGCGCGGTCTGCTGGGCGAAGTAGATCTCGTTTCAAACTCG 517

Qy 503 AGAGCGGCGCCACTACTGTTGGAGTTCAAGTTCATCTACATGGCCAGAGAGCCGCTGC 562
Db 518 AGTCAGGAATCTACTCTGCGCACATGAAGACGTTTACAGATCCAAAGGTGGAGTGA 577

Qy 563 AGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACATCACTCCCAACACGAGACT 622
Db 578 AAGAATTCGCGAATATCACTTTATCCATCATGCTCTGGAGAAACCTACGTGGAGAG 637

Qy 623 ACACCATCTGGAGAGTAGTAG 645
Db 638 GAAGCTTCGTGGAAACAACGAG 660
```

RESULT 10

```
US-09-459-956-4
; Sequence 4, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y., Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; TYPE: DNA
; ORGANISM: Zoanthus sp
US-09-459-956-4

Query Match      16.3%; Score 110.4; DB 4; Length 696;
Best Local Similarity 52.5%; Pred. No. 4.1e-13;
Matches 296; Conservative 0; Mismatches 256; Indels 12; Gaps 2;

Qy 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCAGGTTCAAGGTGCGCATGAGGCG 60
Db 1 ATGGCTCATTTCAAGACCGGTCTAAAGAGAAATGCAATGAATACCAATGGAAGGG 60

Qy 61 ACCGTGAACGCCACGAGTTTCGAGATCGAGGCGGAGGCGGCGCCCGCTACGAGGCG 120
Db 61 TGGCTCAACGAGACATAAATTTGTCATCACGGCGGAGGCATTTGATATCCGTTCAAGGG 120

Qy 121 CACACACCGTGAAGTTGAAGTGCACCAAGGCGGCGCCCTGCGCTTCGCTGGGACATC 180
Db 121 AACAGACTATTAATCTGTGTGTGATCGAAGGGGACATTTGCCATTTTCGAAGACATA 180

Qy 181 CTGTCCCCCGAGTTCCAGTAGCGCTCCAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240
Db 181 TTGTCAGCTGGCTTTAAGTACGGAGACAGGATTTTCACTGAATATCCTCAAGACATAGTA 240

Qy 241 GACTACAAAGAGCTGCTCTTCCCGAGGCTTCAAGTGGGAGCGGTGATGAACTTCGAG 300
Db 241 GACTATTTCAAGAACTCGTGTCTCTGGGATATACATGGGGGAGGTCTTTCTTTTGG 300

Qy 301 GACGCGCGGTGGCGACCGGTGACCCAGGACTCTCTCCCTG-----CAGGACGGCTGCTTC 354
Db 301 GATGGAGCAGTCTCATATGCAATGTAGATATAACAGTAGTGTCAAGAAACTGCAATT 360

Qy 355 ATCTAAGAGTGAAGTTTCATCGGCTGAACTTCCCTCCGACGCGCCCGCTGATGCAAG 414
Db 361 TATCATAAGAGCATATTTAATGGAATGAATTTTCTGCTGATGAGACCTGTGATGAAAG 420

Qy 415 AAGACCATGGCTGGGAGGCTCCACCGAG-----CGCCTGTACCCCGCGAGCGGTG 468
Db 421 ATGACACTAATCTGGAGAGCATCTGCGAGAGATCATGCCAGTACCTAGACGGGATA 480

Qy 469 CTGAAGGCGGAGATTCACAAAGGCCCTGAAGCTGAAGCGCGGCGCACTACTCTGTTGGAG 528
Db 481 CTGAAGGGGATGTCTCCATGTACTCTCTCTGAAAGGATGTTGGGGCTTACCGGTGCAG 540

Qy 529 TTCAAGTCCATCTCATGCGCAAG 552
Db 541 TTCGACACAGTTTACAAGCAAG 564

RESULT 11
US-09-214-909-1
; Sequence 1, Application US/09214909
```

Patent No. 6486382
GENERAL INFORMATION:
APPLICANT: GORDON-KAMM, WILLIAM
APPLICANT: PIERCE, DOROTHY A.
APPLICANT: BOWEN, BENJAMIN
APPLICANT: BIDNEY, DENNIS
APPLICANT: ROSS, MARGIT
APPLICANT: SCHELONGE, CHRISTOPHER
APPLICANT: MILLER, MICHAEL D.
APPLICANT: SANDAHL, GARY
APPLICANT: WANG, LIJUAN
TITLE OF INVENTION: USE OF THE GREEN FLUORESCENT PROTEIN AS A SCREENABLE MARKER FOR PLANT TRANSFORMATION
FILE REFERENCE: 033229/0682
CURRENT APPLICATION NUMBER: US/09/214,909
CURRENT FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US97/07688
PRIOR FILING DATE: 1997-05-01
PRIOR APPLICATION NUMBER: 60/016,345
PRIOR FILING DATE: 1996-05-01
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 717
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
OTHER INFORMATION: encoding GFP
NAME/KEY: CDS
LOCATION: (1)..(714)
US-09-214-909-1

Query Match 14.8%; Score 100.2; DB 4; Length 717;
Best Local Similarity 50.1%; Pred. No. 3.7e-11;
Matches 336; Conservative 0; Mismatches 323; Indels 12; Gaps 3;

QY 1 ATGGCTCTCTCGAGAACGTCATCAGGATTCATCGGCTTCAAGTGGCGATGAGGCG 60
Db 1 ATGTCAAGGGCGAGAGAGCTCTTCAACGGCGTGTGCCATCTCTGTGAGTCTGACGCG 60
QY 61 ACCGTGAAGCGGCACAGATTCGAGATCGAGGGGCGAGGGCGGCCCTTACGAGGCG 120
Db 61 GAGGTGAAGCGGCACAGATTCGAGATCGAGGGGCGAGGGCGGCCCTTACGAGGCG 120
QY 121 CACAACACCGGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCCGCTTGGGACATC 180
Db 121 AAGTCAACCGGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCCGCTTGGGACATC 177
QY 181 CTGTCCCCCAGTTCAGTACGAGTCTCAAGTGTGAGTGAAGCAGCCCGCGACAT- 236
Db 178 CTGTGACCACTTCTCTACGCGTGCAGTGTCTTCCAGGTACCCCGACCAATGAAG 237
QY 237 -CCCCGACTACAGAGGCTCTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGAAGAAC 294
Db 238 CAGCAGACTTCTCAAGTACGAGTCTCCGAGGGCTACGTGAGGAGGAGCAATCTTC 297
QY 295 TTCCAGGACCGCGCGTGGCGAGCGGTGACCCAGATCTCTCTCCCTGAGGAGCGCTCTTC 354
Db 298 TTCAAGGACCGCGCAACTCAAGACCGAGCGCGAGGTGAAGTTCAAGGCGCACCCCTC 357
QY 355 ATCTCAAGGTGAAGTTCACTCGGCTGAATCTCCCTCCGACGCGCCGCTGATGAGAAG 414
Db 358 GTGAACAGGATTTAGCTCAAGGGATCGATTTCAAGGAGGAGCGCAACATCTTCGCCAC 417
QY 415 AAGACCATGGGCTGGGAGGCTTCCACCGAGCGCTGTACCCCGCAGCGCGTGTGAAG 474
Db 418 AAGCTCAGTACAACTCAACTCCCAAA---CGTGTACATCATGCGCGCAAGAGAAG 474
QY 475 GGCAGATCCCAAGGCGCTGAAGTGAAGGACGGCGGCGCACTACCTGTGTGAGTTCAAG 534
Db 475 AACGGCATGAGGTGAATTTCAAGATCAGGCACAAATCAGGAGCGCTCAGTGCAGCTC 534

QY 535 TCCATCTTACATGGCCCAAGAGCCCGTGTAGCTGCCGGCTACTACTACGTGGACACCAAG 594
Db 535 GCTGACCACTACCAGCAGAACACCCCATCGGCGAGGCGCCCGTGTCTCTCCCGCAAC 594
QY 595 CTGACATACCTCCACACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAG 654
Db 595 CACTACCTTCCACCCAGTCCGCCCTCTCCAGGACCCCAACGAGAGGAGGACCATG 654
QY 655 GGCGGCCACCA 665
Db 655 GTGCTCTCTCA 665

RESULT 12
US-09-459-956-3
Sequence 3, Application US/09459956
Patent No. 6342379
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Gonzalez, Ili, Jesus E.
TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
TITLE OF INVENTION: OPTICAL METHODS
FILE REFERENCE: REGEN1290-4
CURRENT APPLICATION NUMBER: US/09/459,956
CURRENT FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: 08/765,860
PRIOR FILING DATE: 1999-05-08
PRIOR APPLICATION NUMBER: 08/481,977
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: PCT/US96/09652
PRIOR FILING DATE: 1996-06-06
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 696
TYPE: DNA
ORGANISM: Zoanthus sp
US-09-459-956-3

Query Match 14.6%; Score 99.2; DB 4; Length 696;
Best Local Similarity 51.2%; Pred. No. 5.8e-11;
Matches 289; Conservative 0; Mismatches 263; Indels 12; Gaps 2;

QY 1 ATGGCTCTCTCGAGAACGTCATCAGGATTCATGCGCTTCAAGGTGCGCATGAGGCG 60
Db 1 ATGGCTCAGTCAAGCAGGCTCTAACAAAGAAATGACATGAAATACCGTATGGAAGG 60
QY 61 ACCGTGAAGCGGCACAGTTCGAGATCGAGGGCGAGGGCGGCCCTTACGAGGCG 120
Db 61 TCGTCGATGACATAAATTTGTGATCAGGAGGAGGCGCATTTGATATTCCTTCAAGGG 120
QY 121 CACACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCCGCTTGGGACATC 180
Db 121 AAACAGGCTATTATCTGTGTGTGTCGAGGTGACCATTTGCCATTGCCGAGACATA 180
QY 181 CTGTCCCCCAGTTCAGTACGCTTCCAGGTGTAAGTGAAGACACCCCGCGACATCCCC 240
Db 181 TTGTGAGTCTCTTAACTACGAAACAGGGTTTCACTGAATATCTCTCAAGACATGTT 240
QY 241 GACTACAGAGCTCTCTTCCCGAGGCTTCAAGTGGGAGCGGTGATGAATCTCGAG 300
Db 241 GACTATTTCAAGAACTCGTGTCTCTGATATATCATGGACAGGTCTTTCTCTTTGAG 300
QY 301 GAGCGGGCGGTGGCGACCGTGACCCAGGACTCTCTCCCTG-----CAGGACGGTGTTC 354
Db 301 GATGAGCAGTTTGCATATGTAATGACAGTGAAGTGTGAAGAAACTGCATG 360
QY 355 ATCTCAAGGTGAAGTTCAATCGGCTGAATCTCCCTCGAGCGCGCGTGTGACAGAG 414
Db 361 TATCATGATCCAAATTTTATGAGTGAATTTTCTGTGTGATGAGACCTGTGTGATGAAAAG 420
QY 415 AAGACCATGGGCTGGAGGCGCTTCCACCGAG-----CGCTGTACCCCGCGACGCGCTG 468

Db 421 ATGACAGATAACTGGAGCCATCTCGAGAGATCATACCAGTACCTAAGACGGGATA 480
Qy 469 CTGAAGGGCGAGATCCACAAGGCCCTGAAGCTGAAGGACGGCGCCACTACCTGGTGAG 528
Db 481 TTGAAGGGGATGCTCCATGATCTCTCTCTGAAGATGTGGCGCTTACGGTGCCAA 540
Qy 529 TTCAAGTCCATCTACATGGCCAAAG 552
Db 541 TTCGACACAGTTTACAAAGCAAG 564

RESULT 13
US-09-486-241-31
; Sequence 31, Application US/09486241
; Patent No. 6472184
; GENERAL INFORMATION:
; APPLICANT: Hegemann, Peter
; TITLE OF INVENTION: METHOD FOR PRODUCING NUCLEIC ACID
; FILE REFERENCE: 3910/06706
; CURRENT APPLICATION NUMBER: US/09/486,241
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/EP98/05219
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: DE19736591.4
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Modified gene from Aquorea victorea
US-09-486-241-31

Query Match 14.6%; Score 99; DB 4; Length 717;
Best Local Similarity 51.0%; Pred. No. 5.4e-11;
Matches 327; Conservative 0; Mismatches 290; Indels 24; Gaps 3;

Qy 1 ATGCGCTCTCCGAGACGTCATACCGAGTTTCATCGCTTCAAGTGGCGCATGGAGGC 60
Db 1 ATGTCCAAAGGGCGAGGAGCTGTTACCGGTGTGTTCCCATCTGCTGGAGCTGGACGGC 60
Qy 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGCGAGGGCGGCGCCCTACGAGGC 120
Db 61 GACGTGAACGCCACAAGTTCTCGGTCTCGCGGAGGGTGAAGGTGACGCCACCTACGGC 120
Qy 121 CACAACACCGTGAAGTTGAAGTGAACCAAGGGCGGCGCCCTGCGCTTGGCGGACATC 180
Db 121 AAGTGACCTGAACTGATCTGACCA---CCGGCAAGCTGCGCCCTGGCGCCACC 177
Qy 181 CTGTCCCCCGAGTTCAGTACGCTCCAGGTGTAGTGAAGCACCACCCCGCGCAT--- 236
Db 178 CTGTGTACCACTTCACTACGCTGTGAGTGTCTTCCCGGTACCCCGACCATGAAG 237
Qy 237 -CCCCGACTACAAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAAC 294
Db 238 CAGCAGCACTTCTCAAGTCCGCCATGCGGAGGGTACGTGAGAGCGCACCATCTTC 297
Qy 295 TTGAGGACCGGCGGTGGCGGACCGTGAACCCAGGATCTCTCCCTGAGGACGGCTGCTTC 354
Db 298 TTCAAGGACACCGGCACTACAAGACCCGCGCGAGGTCAAGTTCAGGAGCGCACCCCTG 357
Qy 355 ATCTACAAGGTGAAGTTCATCGCGGTGAATCTCCCTCCGAGCGCCCGTGTGATGAGAG 414
Db 358 GTGACCGCATCGAGTGAAGGSCATCGACTTCAAGAGAGACGGCAACATCTGGCCAC 417
Qy 415 AAGACCATGGCTGGGAGGCTCCACCGAGCGGCTGTACCCCGGACCGCGTGTGAGAG 474
Db 418 AAGCTGGAGTACAACTACAACTCCCAACACGTTATCATCATGCGCCGACAGAGAAGAC 477
Qy 475 GCGGAGATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCACTACCTGCTGGAGTT--- 530

Db 478 GCATCAAGGTGAATCTCAAGATCCGCCACAACATCGAGGACGGCTCCGTGACGTGGCC 537
Qy 531 -----CAAGTCCATCTAGTGGCCAGAGAGCCCGTGCAGTCCCGCGGTACTAC 579
Db 538 GACCACTACAGCAACACCCCATCGGCGATGGCCCGCTGCTGCTGCTGCTGCTGCTGCT 597
Qy 580 TACGTGGACACCAAGCTGACATCACCTCCCAACACGAGGA 620
Db 598 TACCTGTCCACCCAGTCCGCGCTGTCCAAGGACCCCAACGA 638

RESULT 14
US-09-609-161B-15
; Sequence 15, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUNE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla mulleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-15

Query Match 14.5%; Score 98.4; DB 4; Length 1079;
Best Local Similarity 50.4%; Pred. No. 8.5e-11;
Matches 240; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

Qy 28 GAGTTCATGCGCTTCAAGTGGCGATCGAGGGCACCGTGAAGCCGACGAGTTCGAGATC 87
Db 295 GAAGTATGTCGTATAAGTAATCTGGAAGGAATTTGTAACCAACCATGTTTACAAATG 354
Qy 88 GAGGGCGAGGGCGGCGCCCTACGAGGGGCCACAACACCGGTGAAGTGAAGGTGACC 147
Db 355 GAGGTTTCGGCAAGGGAATATTTTATTCGCAATCACTGGTTTCAGATTCGTGTCAAG 414
Qy 148 AAGGGCGCGCCCTGCGCTGGGACATCTGTCGCCCGGAGTTCAGTACGAGTCC 207
Db 415 AAGGGCGCGCCCTGCGCTGGGACATCTGTCGCCCGGAGTTCAGTACGAGTCC 474
Qy 208 AAGGTGACGTGAAGACACCGCGCATCCCGACTACAAGAGTGTCTCTCCCGAG 267
Db 475 CGTACTTTCAGGAATATCCGAATGATATCAGATTTATTTATACATCATTTCCAGCA 534
Qy 268 GGCTTCAAGTGGGAGCGGTGATGAATTCGAGAGCGCGCGGTGGGACCGTGACCCAG 327
Db 535 GGATTTATGATGAACGAACATTACGTTACGAGATGGCGGACTTGTGAAATTCGTTCA 594
Qy 328 GACTCTCCCTCGAGGACGGCTGCTTCATCTCAAGGTGAAGTTCATCGCGGTGAATTC 387
Db 595 GATATAAATTTAATAGAGACAAGTTCGTCTACAGGTGAATACAAGGTAGTACTTC 654

389 CCCTCCGACGGCCCGTGATGACGAAGAAGACCATGGCTGGGAGGCTCCACCGAGCGC 447
655 CCAGATGATGTCCCGTCATGACAAGACTATCTTAGAATAGAGCCTTCATTGAAGCC 714
448 CTGTACCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCGCTCGAAGCTGAA 503
715 ATGTACATGAATAATGGCGTCTGGTCGGCGAAGTAATCTTGTCATAAATACTAAA 770

```

RESULT 15
US-09-626-581D-64
; Sequence 64, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A-86900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
US-09-626-581D-64

```

Query Match	14.5%	Score 98.4	DB 4	Length 1079
Best Local Similarity	50.4%	Pred. No. 8.5e-11		
Matches 240	Conservative 0	Mismatches 236	Indels 0	Gaps 0
QY	28	GAGTTCATGCGCTTCAAGGTGCGCATGGAGGCGACCGTGAACGGCCACAGTTCAGAGTCC	87	
DB	295	GAAGTAAATGTCGTATAAGTAAATCTGGAAGGAATGTAAACAACCATGTTTTTACAATG	354	
QY	88	GAGGCGGAGGGGAGGGCGGCCCTCAGCAGGGCCACAACACCGTGAAGTTGAAGTGAACC	147	
DB	355	GAGGGTTTCGGCAAGGGAATATTTTATTTCGGCAATCAACTGGTTCAGAAATTCGTGTCA	414	
QY	148	RAAGGCGGCCCCCTGCGCTTCGCTCGGACATCTGTCCCGCCAGTTCACGATACGGCTCC	207	
DB	415	RAAGGGGCCCCCACTCTTTTGCAATTTGATATTGTGTCAACGACCTTTTCAATATGGCAAC	474	
QY	208	RAAGTGTACTGTGAAGACACCCGCGCATCCCGGACTACAAGAAAGTGTCTCTTCCCGGAG	267	
DB	475	CGTACTTTTCAGAAATATCCGAATGATATATCATGATTTATTTTATACAATCATTTTCCAGCA	534	
QY	268	GGCTTCAAGTGGGAGCGGTGATGAATCTGAGAGAGCGGCGGTGGCGACCGTGAACCCAG	327	
DB	535	GAAATTAATGATGAACGAAACATTACGTTACGAAGATGGCGGACTTGTTCGAATTCGTTCAC	594	
QY	328	GACTCCTCCCTCAGGACGGTGTCTTCATCTACAAGGTGAAGTTATCCGGCGTGAACATTC	387	
DB	595	GATATAAAATTAATAGAAAGACAAGTTCTGTACAGAGTGGAAATACAAGAGTATGTAACATTC	654	
QY	388	CCCTCCGAGCGGCCGCTGATGTCAGAAAGACCATGGCTGGGAGGCGCTCCACCGAGCGC	447	
DB	655	CCAGATGATGGTCCCGTCAATGAGAAAGACTATCTTAGGAATAGAGCGCTTCATTTGAAGCC	714	
QY	448	CTGTACCCCGCGACGGCGTCTGAAGGGGAGATCCACAAGGCCCTGAAGCTGAA	503	
DB	715	ATGTGATCATGAATAATGGGGCTCTGGTGGCGGAAGTAACTCTGTATATAAATGAA	770	

Blank Sheet U.S. PTD

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 14:04:25 ; Search time 492.957 Seconds
(without alignments)

6743.643 Million cell updates/sec

Title: US-10-081-864-14

Perfect score: 678

Sequence: 1 atggctctctcgagaacgt.....gacacacctgttctgttaa 678

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	678	100.0	678	14	US-10-006-922-36
2	678	100.0	678	15	US-10-081-864-14
3	676.4	99.8	678	15	US-10-315-920-5
4	672.2	99.1	1050	14	US-10-060-857-7
5	670.2	98.8	675	14	US-10-006-922-38
6	670.2	98.8	675	15	US-10-081-864-13
7	668.4	98.6	678	15	US-10-315-920-3
8	665.2	98.1	678	15	US-10-081-864-7
9	665.2	98.1	678	15	US-10-315-920-1
10	661.6	97.6	681	15	US-10-121-258-3
11	661.6	97.6	681	15	US-10-121-258-23
12	661.6	97.6	4632	15	US-10-161-403-23
13	661.6	97.6	4632	17	US-10-433-640-16
14	661.6	97.6	6984	14	US-10-001-189-45

Sequence 35, Appl
Sequence 37, Appl
Sequence 51, Appl
Sequence 75, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 5, Appl
Sequence 15, Appl
Sequence 9, Appl
Sequence 43, Appl
Sequence 11, Appl
Sequence 1040, Ap
Sequence 1022, Ap
Sequence 1036, Ap
Sequence 6, Appl
Sequence 3, Appl
Sequence 11, Appl
Sequence 2, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 12, Appl
Sequence 3, Appl
Sequence 22, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 45, Appl

US-10-006-922-35
US-10-006-922-37
US-10-214-932-51
US-10-214-932-75
US-10-152-296-1
US-10-739-656-1
US-10-121-258-5
US-10-121-258-7
US-10-121-258-9
US-10-121-258-9
US-10-006-922-43
US-10-081-864-11
US-10-724-178-1040
US-10-724-178-1022
US-10-724-178-1036
US-09-967-772-6
US-10-132-067-3
US-10-006-922-11
US-10-121-258-2
US-10-335-517-6
US-10-334-288-6
US-09-999-745-66
US-09-866-538-11
US-09-794-308-11
US-09-865-291-11
US-10-433-640-12
US-09-797-496B-3
US-10-332-733-22
US-10-314-936-1
US-10-314-936-3
US-10-006-922-45

ALIGNMENTS

RESULT 1

US-10-006-922-36
; Sequence 36, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Mazs, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006.922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma species
US-10-006-922-36

Query Match 100.0%; Score 678; DB 14; Length 678;
Best Local Similarity 100.0%; Pred. No. 4.4e-163;

Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTATGCGCTTCAAGGTGCGCATGGAGGC 60
Db 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTATGCGCTTCAAGGTGCGCATGGAGGC 60

Qy 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGGAGGGGCGCGCCCTTACGAGGGC 120
Db 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGGAGGGGCGCGCCCTTACGAGGGC 120

Qy 121 CACAAACCGGTGAAGTTCGAGATCGAGGGGAGGGGCGCGCCCTTACGAGGGC 180
Db 121 CACAAACCGGTGAAGTTCGAGATCGAGGGGAGGGGCGCGCCCTTACGAGGGC 180

Qy 181 CTGTCCCCCAGTTCAGTTCAGGTCACCAAGGTGAGTGAAGCACCCCGCGCATCCCC 240
Db 181 CTGTCCCCCAGTTCAGTTCAGGTCACCAAGGTGAGTGAAGCACCCCGCGCATCCCC 240

Qy 241 GACTACAGAGTCTCTTCCCGAGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 300
Db 241 GACTACAGAGTCTCTTCCCGAGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 300

Qy 301 GACGGCGGCTGGCGAGCCGTCAGCCAGGCTCTTCCCTGAGGAGCGGTGATGAACCTTCGAG 360
Db 301 GACGGCGGCTGGCGAGCCGTCAGCCAGGCTCTTCCCTGAGGAGCGGTGATGAACCTTCGAG 360

Qy 361 AAGTGAAGTTCATCGGCGTGAATTCCTCCGAGCGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 420
Db 361 AAGTGAAGTTCATCGGCGTGAATTCCTCCGAGCGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 420

Qy 421 ATGGGCTGGGAGGCTTCCACGAGCGCTGTACCCCGAGCGGCTGTGAGGGCGAG 480
Db 421 ATGGGCTGGGAGGCTTCCACGAGCGCTGTACCCCGAGCGGCTGTGAGGGCGAG 480

Qy 481 ATCCAAAGGCTTGAAGTGAAGGAGCGGCGCATCTGCTGTGAGTTCAGTCCATC 540
Db 481 ATCCAAAGGCTTGAAGTGAAGGAGCGGCGCATCTGCTGTGAGTTCAGTCCATC 540

Qy 541 TACATGCCAAGAGCCGTCGAGTGCAGCTGCCCGGTACTACTACGTGACACCAAGCTGGAC 600
Db 541 TACATGCCAAGAGCCGTCGAGTGCAGCTGCCCGGTACTACTACGTGACACCAAGCTGGAC 600

Qy 601 ATCACTTCCCAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGC 660
Db 601 ATCACTTCCCAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGC 660

Qy 661 CACCACCTGTCTCTGTAA 678
Db 661 CACCACCTGTCTCTGTAA 678

RESULT 2
US-10-081-864-14
; Sequence 14, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexander
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR FILING DATE: 2001-12-04, 922
; PRIOR FILING DATE: 2001-12-04, 983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14

LENGTH: 678
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: non-aggregating mutant
US-10-081-864-14

Query Match 100.0%; Score 678; DB 15; Length 678;
Best Local Similarity 100.0%; Pred. No. 4, 4e-163;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTATGCGCTTCAAGGTGCGCATGGAGGC 60
Db 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTATGCGCTTCAAGGTGCGCATGGAGGC 60

Qy 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGGAGGGGCGCGCCCTTACGAGGGC 120
Db 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGGAGGGGCGCGCCCTTACGAGGGC 120

Qy 121 CACAAACCGGTGAAGTTCGAGATCGAGGGGAGGGGCGCGCCCTTACGAGGGC 180
Db 121 CACAAACCGGTGAAGTTCGAGATCGAGGGGAGGGGCGCGCCCTTACGAGGGC 180

Qy 181 CTGTCCCCCAGTTCAGTTCAGGTCACCAAGGTGAGTGAAGCACCCCGCGCATCCCC 240
Db 181 CTGTCCCCCAGTTCAGTTCAGGTCACCAAGGTGAGTGAAGCACCCCGCGCATCCCC 240

Qy 241 GACTACAGAGTCTCTTCCCGAGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 300
Db 241 GACTACAGAGTCTCTTCCCGAGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 300

Qy 301 GACGGCGGCTGGCGAGCCGTCAGCCAGGCTCTTCCCTGAGGAGCGGTGATGAACCTTCGAG 360
Db 301 GACGGCGGCTGGCGAGCCGTCAGCCAGGCTCTTCCCTGAGGAGCGGTGATGAACCTTCGAG 360

Qy 361 AAGTGAAGTTCATCGGCGTGAATTCCTCCGAGCGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 420
Db 361 AAGTGAAGTTCATCGGCGTGAATTCCTCCGAGCGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG 420

Qy 421 ATGGGCTGGGAGGCTTCCACGAGCGCTGTACCCCGAGCGGCTGTGAGGGCGAG 480
Db 421 ATGGGCTGGGAGGCTTCCACGAGCGCTGTACCCCGAGCGGCTGTGAGGGCGAG 480

Qy 481 ATCCAAAGGCTTGAAGTGAAGGAGCGGCGCATCTGCTGTGAGTTCAGTCCATC 540
Db 481 ATCCAAAGGCTTGAAGTGAAGGAGCGGCGCATCTGCTGTGAGTTCAGTCCATC 540

Qy 541 TACATGCCAAGAGCCGTCGAGTGCAGCTGCCCGGTACTACTACGTGACACCAAGCTGGAC 600
Db 541 TACATGCCAAGAGCCGTCGAGTGCAGCTGCCCGGTACTACTACGTGACACCAAGCTGGAC 600

Qy 601 ATCACTTCCCAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGC 660
Db 601 ATCACTTCCCAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGC 660

Qy 661 CACCACCTGTCTCTGTAA 678
Db 661 CACCACCTGTCTCTGTAA 678

RESULT 3
US-10-315-920-5
; Sequence 5, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; FILE REFERENCE: CLON-077CIP
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607

; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-5

Query Match 99.8%; Score 676.4; DB 15; Length 678;
Best Local Similarity 99.9%; Pred. No. 1.1e-162;
Matches 677; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC 60
Db 1 ATGGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC 60
QY 61 ACCGTGAACGGCCACAGTTTCGAGATCGAGGGGCGAGGGCGGGCCCTTACGAGGC 120
Db 61 ACCGTGAACGGCCACAGTTTCGAGATCGAGGGGCGAGGGCGGGCCCTTACGAGGC 120
QY 121 CACAACACCGTGAAGTTGAAGTGACCAAGGGGGCCCTTGCCTTGCCTTGGACATC 180
Db 121 CACAACACCGTGAAGTTGAAGTGACCAAGGGGGCCCTTGCCTTGGACATC 180
QY 181 CTGTCCCTCCAGTTCAGTACGGCTCCAAAGGTGTCAGTGAAGACCCCGCCGACATCC 240
Db 181 CTGTCCCTCCAGTTCAGTACGGCTCCAAAGGTGTCAGTGAAGACCCCGCCGACATCC 240
QY 241 GACTACAGAAGTGTCCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACATTCGAG 300
Db 241 GACTACAGAAGTGTCCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACATTCGAG 300
QY 301 GACGGGGGTGGCGACCGTGAACCCAGAGTCTCTCCCTGCAGAGCGGTGCTTCTATAC 360
Db 301 GACGGGGGTGGCGACCGTGAACCCAGAGTCTCTCCCTGCAGAGCGGTGCTTCTATAC 360
QY 361 AAGTGAAGTTCTACGGGTGAACCTCCCTCGAGGGCCCGTGTATGAGAGAGACC 420
Db 361 AAGTGAAGTTCTACGGGTGAACCTCCCTCGAGGGCCCGTGTATGAGAGAGACC 420
QY 421 ATGGGTGGAGGCTTCAACCGAGCGCTGTACCCCGGAGCGCGTGTGAAGGGCGAG 480
Db 421 ATGGGTGGAGGCTTCAACCGAGCGCTGTACCCCGGAGCGCGTGTGAAGGGCGAG 480
QY 481 ATCCACAAGCCCTGAAGTGAAGGACGGCGGCCACTACCTGGTGAAGTTCAAGTCCATC 540
Db 481 ATCCACAAGCCCTGAAGTGAAGGACGGCGGCCACTACCTGGTGAAGTTCAAGTCCATC 540
QY 541 TACATGCCCAAGAGCCCTGACAGTCCCGGCTACTACTAGTGGACACCAAGTGGAC 600
Db 541 TACATGCCCAAGAGCCCTGACAGTCCCGGCTACTACTAGTGGACACCAAGTGGAC 600
QY 601 ATCACTCCCAACAGAGACTACACCATCTGTGGAGCAGTACGAGCGACCGAGGGCCGC 660
Db 601 ATCACTCCCAACAGAGACTACACCATCTGTGGAGCAGTACGAGCGACCGAGGGCCGC 660
QY 661 CACCACCTGTTCTCTGTA 678
Db 661 CACCACCTGTTCTCTGTA 678

RESULT 4
US-10-060-857-7
; Sequence 7, Application US/10060857
; Publication No. US20020132318A1
; GENERAL INFORMATION:
; APPLICANT: Anticancer, Inc.
; APPLICANT: Zhao, Ming

; APPLICANT: Jiang, Ping
; APPLICANT: Xu, Mingxu
; APPLICANT: Yang, Meng
; TITLE OF INVENTION: IMPROVED FLUORESCENT PROTEINS
; FILE REFERENCE: 31276-20032.00
; CURRENT APPLICATION NUMBER: US/10/060,857
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US 60/264,932
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1050
; TYPE: DNA
; ORGANISM: Coral
US-10-060-857-7

Query Match 99.1%; Score 672.2; DB 14; Length 1050;
Best Local Similarity 99.6%; Pred. No. 1.4e-161;
Matches 674; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC 60
Db 289 ATGGCCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCGATGGAGGC 348
QY 61 ACCGTGAACGGCCACAGTTTCGAGATCGAGGGGCGAGGGCGGGCCCTTACGAGGC 120
Db 349 ACCGTGAACGGCCACAGTTTCGAGATCGAGGGGCGAGGGCGGGCCCTTACGAGGC 408
QY 121 CACAACACCGTGAAGTTGAAGTGACCAAGGGGGCCCTTGCCTTGCCTTGGGACATC 180
Db 409 CACAACACCGTGAAGTTGAAGTGACCAAGGGGGCCCTTGCCTTGCCTTGGGACATC 468
QY 181 CTGTCCCTCCAGTTCAGTACGGCTTCAAGTGAAGACCCCGCCGACATCC 240
Db 469 CTGTCCCTCCAGTTCAGTACGGCTTCAAGTGAAGACCCCGCCGACATCC 528
QY 241 GACTACAGAAGTGTCTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACATTCGAG 300
Db 529 GACTACAGAAGTGTCTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACATTCGAG 588
QY 301 GACGGGGGTGGCGACCGTGAACCCAGAGTCTCTCCCTGCAGAGCGGTGCTTCTATAC 360
Db 589 GACGGGGGTGGCGACCGTGAACCCAGAGTCTCTCCCTGCAGAGCGGTGCTTCTATAC 648
QY 361 AAGTGAAGTTCTACGGGTGAACCTTCCCTCGAGCGCCCGTGTATGAGAGAGACC 420
Db 649 AAGTGAAGTTCTACGGGTGAACCTTCCCTCGAGCGCCCGTGTATGAGAGAGACC 708
QY 421 ATGGGTGGAGGCTTCAACCGAGCGCTGTACCCCGGAGCGCGTGTGAAGGGCGAG 480
Db 709 ATGGGTGGAGGCTTCAACCGAGCGCTGTACCCCGGAGCGCGTGTGAAGGGCGAG 768
QY 481 ATCCACAAGCCCTGAAGTGAAGGACGGGGCCACTACCTGGTGAAGTTCAAGTCCATC 540
Db 769 ATCCACAAGCCCTGAAGTGAAGGACGGGGCCACTACCTGGTGAAGTTCAAGTCCATC 828
QY 541 TACATGCCCAAGAGCCCGTGAAGTCCCGGCTACTACTAGTGGGACACCAAGTGGAC 600
Db 829 TACATGCCCAAGAGCCCGTGAAGTCCCGGCTACTACTAGTGGGACACCAAGTGGAC 888
QY 601 ATCACTCCCAACAGAGACTACACCATCTGTGGAGCAGTACGAGCGACCGAGGGCCGC 660
Db 889 ATCACTCCCAACAGAGACTACACCATCTGTGGAGCAGTACGAGCGACCGAGGGCCGC 948
QY 661 CACCACCTGTTCTCTGTA 677
Db 949 CACCACCTGTTCTCTGTA 965

RESULT 5
US-10-006-922-38
; Sequence 38, Application US/10006922

QY 421 ATGGGTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCGTGTGTAAGGGCGAG 480
Db 421 ATGGGTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGGCGTGTGTAAGGGCGAG 480
QY 481 ATCCACAAGGCGCTGAAGCTGAAGGACGGCGCACTACCTGCTGAGTTCAAGTCCATC 540
Db 481 ACCACAAGGCGCTGAAGCTGAAGGACGGCGCACTACCTGCTGAGTTCAAGTCCATC 540
QY 541 TACATGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
Db 541 TACATGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
QY 601 ATCACTCCCAACAGAGACTACACCATCGTGGAGCAGTACGAGGCAACCGAGGCGGC 660
Db 601 ATCACTCCCAACAGAGACTACACCATCGTGGAGCAGTACGAGGCAACCGAGGCGGC 660
QY 661 CACCACCTGTTCTCTG 675
Db 661 CACCACCTGTTCTCTG 675

RESULT 7

US-10-315-920-3
; Sequence 3, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Pradkov, Arcady Fedorovich
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-3

Query Match 98.6%; Score 668.4; DB 15; Length 678;
Best Local Similarity 99.1%; Pred. No. 1.2e-160;
Matches 672; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGCGCATGGAGGC 60
Db 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGCGCATGGAGGC 60
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTTACGAGGC 120
Db 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTTACGAGGC 120
QY 121 CACAACACCGTGAAGTTGAAGTGAACAAGGGGGCGCCCTTCCGCTTGGGACATC 180
Db 121 CACAACACCGTGAAGTTGAAGTGAACAAGGGGGCGCCCTTCCGCTTGGGACATC 180
QY 181 CTGTCCCGCCAGTTCAGTACGGCTCCAAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240
Db 181 CTGTCCCGCCAGTTCAGTACGGCTCCAAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240
QY 241 GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGCGTGTAGACTTCGAG 300
Db 241 GACTACAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGGCGTGTAGACTTCGAG 300
QY 301 GACGGCGGTGGCGACCGTGTACCCAGGACTCTCTCCCTGACGACGGCTTTCATCTAC 360

Db 301 GACGGCGGTGGCGACCGTGAACCCAGGATCTCTCCCTCGAGGACGGCTTTCATCTAC 360
QY 361 AAGGTGAAGTTTCATCGCGTGAACCTTCCCTCCGACGGCGCCCGTGTATGCAAGAGACC 420
Db 361 AAGGTGAAGTTTCATCGCGTGAACCTTCCCTCCGACGGCGCCCGTGTATGCAAGAGACC 420
QY 421 ATGGGCTGGAGGCTTCACCGAGCGCTGTACCCCGCGACGGCGTGTGTAAGGGCGAG 480
Db 421 ATGGGCTGGAGGCTTCACCGAGCGCTGTACCCCGCGACGGCGTGTGTAAGGGCGAG 480
QY 481 ATCCACAAGGCGCTGAAGCTGAAGGACGGGGCCACTTACCTGGTGGAGTTCAAGTCCATC 540
Db 481 ATCCACAAGGCGCTGAAGCTGAAGGACGGGGCCACTTACCTGGTGGAGTTCAAGTCCATC 540
QY 541 TACATGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
Db 541 TACATGCCAAGAAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGAC 600
QY 601 ATCACTCCCAACAGAGACTACACCATCGTGGAGCAGTACGAGGCAACCGAGGCGGC 660
Db 601 ATCACTCCCAACAGAGACTACACCATCGTGGAGCAGTACGAGGCAACCGAGGCGGC 660
QY 661 CACCACCTGTTCTCTGTA 678
Db 661 CACCACCTGTTCTCTGTA 678

RESULT 8

US-10-081-864-7
; Sequence 7, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey
; APPLICANT: Lukanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: NO. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-10-081-864-7

Query Match 98.1%; Score 665.2; DB 15; Length 678;
Best Local Similarity 98.8%; Pred. No. 8e-160;
Matches 670; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGCGCATGGAGGC 60
Db 1 ATGGCTCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGCGCATGGAGGC 60
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTTACGAGGC 120
Db 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCTTACGAGGC 120
QY 121 CACAACACCGTGAAGTTGAAGTGAACAAGGGGGCGCCCTTCCGCTTGGGACATC 180
Db 121 CACAACACCGTGAAGTTGAAGTGAACAAGGGGGCGCCCTTCCGCTTGGGACATC 180
QY 181 CTGTCCCGCCAGTTCAGTACGGCTCCAAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240
Db 181 CTGTCCCGCCAGTTCAGTACGGCTCCAAAGGTGTACGTGAAGCACCCCGCGACATCCCC 240

Qy	241	GACTACAAGAACTCTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAATCTTCGAG	300
Db	241	GACTACAAGAACTCTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGATGAATCTTCGAG	300
Qy	301	GACGGGGCGTGGCAGACCGTGACCGAGACCTCCCTGCAGAGACGGCTGTTCACTTAC	360
Db	301	GACGGGGCGTGGTGAACCGTGACCGAGACTTCTCCCTGCAGAGCGCTGTTCACTTAC	360
Qy	361	AAGTGAAGTTTCATCGGCGTGAACTTCCCTCCGACGGCCCCCTGTATGCAGAAAGAAGCC	420
Db	361	AAGTGAAGTTTCATCGGCGTGAACTTCCCTCCGACGGCCCCGTATGCAGAAAGAAGCC	420
Qy	421	ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGCAGCGCGTGTGAAGGGCGAG	480
Db	421	ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGCAGCGCGTGTGAAGGGCGAG	480
Qy	481	ATCCACAAGGCCCTGAAGCTGAAGCAGCGCGGCCACTACCTGTGGAGTTCAAAGTCCATC	540
Db	481	ATCCACAAGGCCCTGAAGCTGAAGCAGCGCGGCCACTACCTGTGGAGTTCAAAGTCCATC	540
Qy	541	TACATGGCCACAAGCGCGTGCAGTGCCTCCGGCTACTACTACGTGGACACCAAGCTGGAC	600
Db	541	TACATGGCCACAAGCGCGTGCAGTGCCTCCGGCTACTACTACGTGGACTCCAAGCTGGAC	600
Qy	601	ATCACCTCCCAAGCAGAGACTTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCCGC	660
Db	601	ATCACCTCCCAAGCAGAGACTTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCCGC	660
Qy	661	CACCACCTGTTCTCTTAA	678
Db	661	CACCACCTGTTCTCTTAA	678

```

RESULT 9
US-10-315-920-1
; Sequence 1, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-1

```

Query Match	98.1%	Score 665.2;	DB 15;	Length 678;
Best Local Similarity	98.8%;	Pred. NO. 8e-160;		
Matches 670; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;

Qy	1	ATGGCTCTCTCGAGAA	CGTCTACACCGAGTT	CATGCGCTCAAGGTGCGCATGAGGGC	60
Db	1	ATGGCTCTCTCAAGAA	CGTCTACACGAGTT	CATGCGTTCAAGGTGCGCATGAGGGC	60
Qy	61	ACCGTGAACGGCCAC	CGAGTTTCAGATCAG	GGCGAGGGCGAGGGCCGCCCTTACGAGGGC	120
Db	61	ACCGTGAACGGCCAC	CGAGTTTCAGATCAG	GGCGAGGGCGAGGGCCGCCCTTACGAGGGC	120
Qy	121	CACAACACCGTGAAG	TTGAAGTGTACCAAG	GGCGGGCGCCCTTCGCCCTTCGGCTCGGACATC	180

5	TGCGCTCTCCAAGAACGTCAATCAAGGAGTTCAATGCGCTTCAAGGTGCGCATGAGGGCA	64
62	CGGTGAACGGCCACGAGTTTCAGATCGAGGCGAGGGCGAGGCCGCCCTTACGAGGGCC	121
65	CGGTGAACGGCCACGAGTTTCAGATCGAGGCGAGGGCGAGGGCCGCCCTTACGAGGGCC	124
122	ACAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTGCCCTTCGCGCTCGGACATCC	181
125	ACAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTGCCCTTCGCGCTCGGACATCC	184
182	TGTCCTCCCGACGTTCCAGTACGGCTCCAAAGGTGTACGTGAAGCACCCCGCCGACATCCCG	241
185	TGTCTCCCGACGTTCCAGTACGGCTCCAAAGGTGTACGTGAAGCACCCCGCCGACATCCCG	244
242	ACTACAAGAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG	301
245	ACTACAAGAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGG	304
302	ACGGGGCGGTGGCGAACCGTGACCCAGGACTCTCTCCCTGCAGGACGGCTGTCTCATCTACA	361
305	ACGGGGCGGTGGTGNACGTGACCCAGACTCTCTCCCTGCAGGACGGCTGTCTCATCTACA	364
362	AGGTGAAGTTTCATCGGCGTGAACTTCCCTTCGACGGCCCCCGTGTGACAGAGAAGACCA	421
365	AGGTGAAGTTTCATCGGCGTGAACTTCCCTTCGACGGCCCCCGTGTGACAGAGAAGACCA	424
422	TGGGCTGGAGGGCTCCACCGAGCGCTGTACCCCGCGACGCGCTGTGAAGGGCGAGA	481
425	TGGGCTGGAGGGCTCCACCGAGCGCTGTACCCCGCGAGCGCTGTGAGAGGGCGAGA	484
482	TCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGTGGAGTTCAAGTCCATCT	541
485	TCCACAAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGTGGAGTTCAAGTCCATCT	544
542	ACATGGCCACAGAGCCGTGACGTGCCCGCTACTACTACGTGGACACCAAGCTGGACA	601
545	ACATGGCCACAGAGCCGTGACGTGCCCGCTACTACTACGTGGACACCAAGCTGGACA	604
602	TCACTCTCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGCGC	661
605	TCACTCTCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGCGC	664
662	ACCACCTGTTCCTGTA	677
665	ACCACCTGTTCCTGTA	680

RESULT 11

```

US-10-121-258-23
; Sequence 23, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CPI
; CURRENT APPLICATION NUMBER: US/10/121.258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: nucleotide sequence encoding DsRed with mammalian
; OTHER INFORMATION: codon usage
US-10-121-258-23

```

RESULT 12

```

US-10-161-403-29
; Sequence 29, Application US/10161403
; Publication No. US20030119104A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shellard, Joan
;
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
;
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10/161,403
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30

```


Publication No. US20020173634A1
GENERAL INFORMATION:
APPLICANT: FRASER JR., MALCOLM J.
APPLICANT: LI, XU
APPLICANT: BEAM, TERESA
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSPOSITION USING
TITLE OF INVENTION: MINIMAL SEGMENTS OF THE EUKARYOTIC TRANSFORMATION
TITLE OF INVENTION: VECTOR PIGGYBAC
FILE REFERENCE: 835910-92098
CURRENT APPLICATION NUMBER: US/10/001,189
CURRENT FILING DATE: 2001-10-30
PRIOR APPLICATION NUMBER: 60/244,984
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/244,677
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 45
LENGTH: 6984
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:
OTHER INFORMATION: PBAP3-DsRed-orf sequence
US-10-001-189-45

Query Match 97.6%; Score 661.6; DB 14; Length 6984;
Best Local Similarity 98.7%; Pred. No. 7.5e-159;
Matches 667; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGCTCTCCGAGAACGTCATCCAGGTTTCATGGCTTCAAGTGGCGCATGGAGGCA 61
DB 2975 TGGCTCTCCGAGAACGTCATCCAGGTTTCATGGCTTCAAGTGGCGCATGGAGGCA 3034

QY 62 CCGTGAACCGCCACAGTTTCGAGATCGAGGGGAGGGGCGGCGCCCTACGAGGGCC 121
DB 3035 CCGTGAACCGCCACAGTTTCGAGATCGAGGGGAGGGGCGGCGCCCTACGAGGGCC 3094

QY 122 ACAACACCGTGAAGTTGAAGTTGAACCAAGGGGCGCCCTCGCTTGGCTGGACATCC 181
DB 3095 ACAACACCGTGAAGTTGAAGTTGAACCAAGGGGCGCCCTCGCTTGGCTGGACATCC 3154

QY 182 TGTGCCCCCAGTTCACAGTACGGTCCCAAGTGTACGTGAAGACACCCCGCGACATCCCG 241
DB 3155 TGTGCCCCCAGTTCACAGTACGGTCCCAAGTGTACGTGAAGACACCCCGCGACATCCCG 3214

QY 242 ACTACAAGAGTGTCTCTCCCGAGGGCTTCAAGTGGAGCGCGGTGATGAATTCGAGG 301
DB 3215 ACTACAAGAGTGTCTCTCCCGAGGGCTTCAAGTGGAGCGCGGTGATGAATTCGAGG 3274

QY 302 ACGGCGCGTGGCGACCGTGAACCGACCTCTCCCTCGAGACGCGCTCTTCATCTACA 361
DB 3275 ACGGCGCGTGGCGACCGTGAACCGACCTCTCCCTCGAGACGCGCTCTTCATCTACA 3334

QY 362 AGGTGAAGTTCATCGGGCGTGAATCTCCCTCCGACGCGCGCGGTGATGACAGAGACCA 421
DB 3335 AGGTGAAGTTCATCGGGCGTGAATCTCCCTCCGACGCGCGCGGTGATGACAGAGACCA 3394

QY 422 TGGGTGGGAGGCTTCAACGAGGCGCTGTACCCCGCGAGCGCGGTGAGAGGGGCGAGA 481
DB 3395 TGGGTGGGAGGCTTCAACGAGGCGCTGTACCCCGCGAGCGCGGTGAGAGGGGCGAGA 3454

QY 482 TCCACAGGCGCTTGAAGCTGAAGAGCGGGCGCACTACCTGTGGAGTTCAAGTCCATCT 541
DB 3455 TCCACAGGCGCTTGAAGCTGAAGAGCGGGCGCACTACCTGTGGAGTTCAAGTCCATCT 3514

QY 542 ACATGGCCAAAGAGCCCGTCAAGTCCCGGCTACTACTACGTGGGACACCAAGCTGGACA 601
DB 3515 ACATGGCCAAAGAGCCCGTCAAGTCCCGGCTACTACTACGTGGGACACCAAGCTGGACA 3574

QY 602 TCACCTCCCAACAGGAGTACACCATCTGTGAGGAGTACGAGCGCACCGAGGGCGCGC 661
DB 3575 TCACCTCCCAACAGGAGTACACCATCTGTGAGGAGTACGAGCGCACCGAGGGCGCGC 3634

QY 662 ACCACCTGTTCCTGTA 677
DB 3635 ACCACCTGTTCCTGTA 3650

RESULT 15
US-10-006-922-35
Sequence 35, Application US/10006922
Publication No. US20020197676A1
GENERAL INFORMATION:
APPLICANT: Lukyanov, Sergey A.
APPLICANT: Fradkov, Arcady F.
APPLICANT: Labas, Yulii A.
APPLICANT: Matz, Mikhail V.
APPLICANT: Tersikh, Alexey
TITLE OF INVENTION: No. US20020197676A1 Chromophores/Fluorophores and
TITLE OF INVENTION: Methods for Using the Same
FILE REFERENCE: CLON-035CIP
CURRENT APPLICATION NUMBER: US/10/006,922
CURRENT FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 09/120,330
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/457,898
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,144
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,477
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/457,556
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/444,338
PRIOR FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 35
LENGTH: 681
TYPE: DNA
ORGANISM: Discosoma species
US-10-006-922-35

Query Match 97.3%; Score 660; DB 14; Length 681;
Best Local Similarity 98.5%; Pred. No. 1.7e-158;
Matches 666; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 TGGCTCTCCGAGAACGTCATCCAGGTTTCATGGCTTCAAGTGGCGCATGGAGGCA 61
DB 5 TGGCTCTCCGAGAACGTCATCCAGGTTTCATGGCTTCAAGTGGCGCATGGAGGCA 64

QY 62 CCGTGAACCGCCACAGTTTCGAGATCGAGGGGAGGGGCGGCGCCCTACGAGGGCC 121
DB 65 CCGTGAACCGCCACAGTTTCGAGATCGAGGGGAGGGGCGGCGCCCTACGAGGGCC 124

QY 122 ACAACACCGTGAAGTTGAAGTTGAACCAAGGGGCGCCCTCGCTTGGCTGGACATCC 181
DB 125 ACAACACCGTGAAGTTGAAGTTGAACCAAGGGGCGCCCTCGCTTGGCTGGACATCC 184

QY 182 TGTGCCCCCAGTTCACAGTACGGTCCCAAGTGTACGTGAAGACACCCCGCGACATCC 241
DB 185 TGTGCCCCCAGTTCACAGTACGGTCCCAAGTGTACGTGAAGACACCCCGCGACATCC 244

QY 242 ACTACAAGAGTGTCTCTCCCGAGGGCTTCAAGTGGAGCGCGGTGATGAATTCGAGG 301
DB 245 ACTACAAGAGTGTCTCTCCCGAGGGCTTCAAGTGGAGCGCGGTGATGAATTCGAGG 304

QY 302 ACGGCGCGTGGCGACCGTGAACCGACGACTCTCCCTCGAGAGCGGTGCTTCATCTACA 361
DB 305 ACGGCGCGTGGCGACCGTGAACCGACGACTCTCCCTCGAGAGCGGTGCTTCATCTACA 364

QY 362 AGGTGAAGTTCATCGGGCGTGAATCTCCCTCGAGAGCGGGCGCGGTGATGAGAGAGACCA 421
DB 365 AGGTGAAGTTCATCGGGCGTGAATCTCCCTCGAGAGCGGGCGCGGTGATGAGAGAGACCA 424

```
QY 422 TGGGTGGAGGCGCTCCACCGAGCGCCTGTATACCCCGCGACGGCGTCTGTAAGGGCGAGA 481
Db |||||||
QY 425 TGGGTGGAGGCGCTCCACCGAGCGCCTGTATACCCCGCGACGGCGTCTGTAAGGGCGAGA 484
Db |||||||
QY 482 TCCACARGGCGCTGAAGCTGAAGGACGGCGGCCCACTACCTGGTGGAGTTCAAGTCCATCT 541
Db |||||||
QY 485 TCCACARGGCGCTGAAGCTGAAGGACGGCGGCCCACTACCTGGTGGAGTTCAAGTCCATCT 544
Db |||||||
QY 542 ACATGGCCCAAGAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACA 601
Db |||||||
QY 545 ACATGGCCCAAGAGCCCGTGCAGCTGCCCGGCTACTACTAGTGGACTCCCAAGCTGGACA 604
Db |||||||
QY 602 TCACCTCCCAACGAGGACTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 661
Db |||||||
QY 605 TCACCTCCCAACGAGGACTACACCATCGTGGAGGAGTACGAGCGCACCGAGGGCGGCC 664
Db |||||||
QY 662 ACCACCTGTTCTGTGTA 677
Db |||||||
QY 665 ACCACCTGTTCTGTGTA 680
Db |||||||
```

Search completed: August 3, 2004, 19:01:41
Job time : 494.957 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 11:02:24 ; Search time 3038.03 Seconds
(without alignments)
6664.376 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 678
Sequence: 1 atggctctctcgagaacgt.....ggcaccacctgtctctgttaa 678

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- EST:*
 - 1: em_estba:*
 - 2: em_estum:*
 - 3: em_estin:*
 - 4: em_estov:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_esti:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rod:*
 - 26: em_gss_phg:*
 - 27: em_gss_vri:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.8	12.1	681	12	BI386888 BFL26_001
2	79.4	11.7	550	13	BO551001 H4006G01-
3	77.8	11.5	629	12	BI384961 BFLG2_003
4	77.6	11.4	707	12	BI382442 BFLG2_000

5	77.4	11.4	561	12	BI379061 BFLG1_000
6	77.4	11.4	577	12	BI386955 BFL26_001
7	77.4	11.4	671	12	BI381780 BFLG1_003
8	77.4	11.4	687	12	BI387917 BFL26_002
9	76.4	11.3	459	12	BI386952 BFL26_001
10	75.8	11.2	661	12	BI385436 BFL26_000
11	75.4	11.1	637	12	BI379468 BFLG1_001
12	75.4	10.7	591	12	BI382373 BFLG2_000
13	69.6	10.3	675	12	BI382638 BFLG2_000
14	69.4	10.2	537	12	BI376909 BFLG3_000
15	69.4	10.2	826	29	CG448047 OSVGS13TH
16	68	10.0	643	12	BI386699 BFL26_001
17	66.4	9.8	670	14	CA258642 SCCRT301
18	65.6	9.7	571	12	BI387406 BFL26_002
19	65.6	9.7	613	12	BI386976 BFL26_001
20	65.2	9.6	629	12	BI388023 BFL26_002
21	64.2	9.5	631	13	CA137844 SCCRT200
22	64.2	9.5	689	12	BI380055 BFLG1_001
23	64.2	9.5	702	14	CA258524 SCCRT300
24	64.2	9.5	801	29	CG445199 OGVHS19TH
25	64	9.4	568	28	BZ708497 OGBAP91TM
26	64	9.4	629	12	BI380008 BFLG1_001
27	64	9.4	850	29	CG210258 OGBXN21TV
28	62	9.1	501	13	BI718163 1031024C0
29	61.8	9.1	560	10	BE497059 WHE0763_H
30	61.8	9.1	725	13	BU672170 WHE3301_D
31	61.4	9.1	683	14	CA166969 SCCSB100
32	61.4	9.1	750	13	BO605927 BRY_1517
33	61.4	9.1	799	12	EG310393 HVSMEC001
34	61.2	9.0	806	29	CC607863 OGLBY87TH
35	61.2	9.0	829	28	BZ643262 OCCQ08TC
36	61.2	9.0	845	13	CA126366 SCSGLR108
37	60.8	9.0	655	14	CD885812 G118-1001
38	60.8	9.0	624	9	AI948344 603043C03
39	60.6	8.9	624	9	AI948344 603043C03
40	60.6	8.9	872	14	CB657500 OSJNEC120
41	60.4	8.9	559	12	BM324075 PIC1_24_H
42	60.4	8.9	607	12	BM323555 PIC1_20_F
43	60.4	8.9	643	12	BU256281 BU256281
44	60.4	8.9	650	12	BU258703 BU258703
45	60.4	8.9	651	12	BU257736 BU257736

ALIGNMENTS

RESULT 1
BI386888
LOCUS
DEFINITION
MPMGp531) Branchiostoma floridae cDNA clone MPMGP53119122 5', mRNA sequence.
ACCESSION
BI386888
VERSION
EST.
KEYWORDS
SOURCE
ORGANISM
Branchiostoma floridae (Florida lancelet)
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.
REFERENCE
1 (bases 1 to 681)
AUTHORS
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J., Herwig, R., Vingron, M. and Lehrach, H.
TITLE
New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes

BI386888 681 bp mRNA linear EST 26-AUG-2003
BFL26_001591 Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531) Branchiostoma floridae cDNA clone MPMGP53119122 5', mRNA sequence.

Handwritten signature

Genomes
Genome Res. 13 (6A), 1056-1066 (2003)

JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Panopoulou G
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inhestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>.
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTTACACTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACGCGCTGGCGGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CGGTCGCGAATTCGCGGT-3' pSport3/86
High quality sequence stop: 681.

FEATURES

Location/Qualifiers
1..681
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMG531119122"
/tissue type="whole embryo"
/dev stage="26 hrs (neurula stage)"
/lab host="E.coli, XL1 blue"
/clone.lib="Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMG531)"
/note="vector: pSport1 (Gibco BRL); Site 1: SalI, XbaI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5')-pGATAGTTCATGATCGGAGCGCGCC (T)15-3' and a SalI 5'-TCGACCCAGCGCTCCG-3' adapters (Gibco BRL)".

ORIGIN

Query Match 12.1%; Score 81.8; DB 12; Length 681;
Best Local Similarity 49.8%; Pred. No. 0.028;
Matches 262; Conservative 0; Mismatches 258; Indels 6; Gaps 2;
Qy 42 CAAGTGGGATGAGGCGACCGTGAACGCGGAGTTCGAGATCGAGGCGGAGCGGA 101
Db 82 CGACCTTCATCTTCGGCTCCGTCATGCGGAGTTCGACTGCTGGAGGCGGAA 141
Qy 102 GGGCGGCGCTAGAGGGGCGACACCGTGAAGTTGAAGGTGACCAAGGCGGCGCCCT 161
Db 142 GGGCAACCGCAACGATGGAACGCTCGAGACCACTGTGAAATCCACCGGGGG-CCCT 198
Qy 162 GCCTTCGCTGGACATCTGTCCTCCCGAGTTCAGTACGCTCCAGGTGTACGTGA 221
Db 199 GCCTTCGCTCCCGTGTGTGATCGGACCAACCTGGGTACGGCTTCTACAGTACCTGCC 258
Qy 222 GCACCCCGCGACATCCCGAGCTACAGAGCTGTCTTCCCGAGGGCTTCAAGTGGGA 281
Db 259 CTTCCCTGGCGGCTCCTACCTTCCAAACCGCCTACGAGGAGGTACAGGTTC 318
Qy 282 CGGGTGTGAATTCGAGACCGGCGGTGGCGACCGGTGACCCAGATCTCTCCCTGCA 341
Db 319 CCGTGTGTCAAGTTTGAAGACCGGCGAGTGTCAATTGCACTTCGCGTACACCTACNA 378
Qy 342 GGACGGTGTCTTCATCTACAAGTGAAGTTTCATCGGCGGTGAATCTTCCCTCCGACGGCC 401
Db 379 GGGCGGCAAGATCAAGGGGAGTTCAGCTGATCGGTCAGGTTCCTTCGCGGGTCC 438
Qy 402 CGTGATG---CAGAAGAGACCATGGGCTGGAGGCGCTCCACGAGGCGCTGTACCCCG 458
Db 439 CGTGATGTCCGGCGGAGTACACACCTGGACAGGAGCGTGGCGAACTCGAGTGTCCGA 498
Qy 459 CGACGGGCTGTCAAGGGCGAGATCCACAAGGCGCTGAAGCTGAAGGAGCGGCGGCACTA 518
Db 499 CGACCGCACCATCACCGGCACTAAGAGTGGAGCTTCTGCACCAACCGATGGGAAGCGCTA 558

Qy 519 CTTGGTGGAGTTCAGTCCATCTACATGCGCCAAAGAGCCCGTGCAG 564
Db 559 TCAGCGGAGCTGCAGAGCACTACACCTTCGCCAAGCGCTCCCG 604

RESULT 2

Q551001
LOCUS BQ551001
DEFINITION H4006G01-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
H4006G01 5', mRNA sequence.
ACCESSION BQ551001
VERSION BQ551001
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 550)
VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin, P.R., Stagg, C.A., Bassey, U., Aiba, K., Hamatani, T., Kargul, G.J., Luo, A.G., Kelsso, J., Hide, W. and Ko, M.S.H.,
Assembly, verification, and initial annotation of NIA 7.4K mouse cDNA clone set
Genome Res. 12 (12), 1999-2003 (2002)
12466305
Other ESTs: H4006G01-3
Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@gsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please visit http://lgsun.grc.nia.nih.gov/cDNA/NIA_7_4k.html for details.
Plate: H4006 row: G column: 01
Seq primer: -21M13 Reverse
High quality sequence stop: 550
POLYA-No.

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

Location/Qualifiers

1..550

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="niaEST:H4006G01-5"

/db_xref="taxon:10090"

/clone="H4006G01"

/sex="mixed"

/dev stage="mixed"

/lab_host="DH10B"

/clone_lib="NIA Mouse 7.4K cDNA Clone Set"

/note="vector: pSPORT1; Site 1: SalI; Site 2: NotI; This clone is among a rearranged set of 7,407 clones from more than 20 cDNA libraries."

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

0; Mismatches

176; Indels

9; Gaps

2;

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Db 329 CCTGACCTACGGCGTGCAGTCTTTCAGCGCTTACCCGACACGCTGAAGCAGCAGCACTT 388
Qy 246 CAAGAAGCTGTCTCTCCCGAGGGTTCAGTGGAGCGCGTGATGAATCTCGAGGAGG 305
Db 389 CTTCAAGTCCGCTCCCGAGGCTACGTCAGGAGCGCACCATCTCTTCTCAAGGAGA 448
Qy 306 CGCGCTGGCGACCGTGAACCCAGCACTCTCTCCCTGAGGAGCGGCTCTTCATCTACAAGT 365
Db 449 CGGCACTACAGACCGCGCGAGGTGAAGTTTCAGGGCGACACCTGCTGTAACCGCAT 508
Qy 366 GAAGTTCATCGCGGTGAATCTCCCTCCGACGCGCCCGT 404
Db 509 CGAGTGAAGGGCATCGACTTCAAGGAGGACGCAACAT 547

RESULT 3
LOCUS
DEFINITION
BFLG2_003138 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498E05107 5',
mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Branchiostoma floridae (Florida lancelet)
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.

REFERENCE
AUTHORS
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
Herwig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes

Genome Res. 13 (6A), 1056-1066 (2003)

JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Panopoulou G
Laboratory 145, Dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers
FORWARD: 5' CCCAGCTTACACTTATGTTCCGGCTCG 3' (M13RSP)
BACKWARD: 5' GGTATTACCCAGTGGGAAAGGGGATG 3' (M13FSP)
Insert length: 1 Std Error: 0.00
Seq primer: 5'-CCGTCGCGAATTCCTCGGT-3' pSport3/86
High quality sequence stop: 629.

FEATURES
source
1..629
Location/Qualifiers
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMPGp498E05107"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name

convention: BFLG or MPMPGp498)"
/notes="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
directionally cloned in pSport1 vector using a NotI
(5'-PACTAGTTCTAGATCGGAGCGCGCCC (T)15-3' and a SalI 5'-
TCGACCCACGCTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 11.5%; Score 77.8; DB 12; Length 629;
Best Local Similarity 51.7%; Pred. No. 0.1;
Matches 203; Conservative 0; Mismatches 187; Indels 3; Gaps 1;
Qy 39 CTTCAAGTGGCATGAGGGGACCGTGAAGCGGCCAGGTTTCGAGATCGAGGCGGAGG 98
Db 104 CTTCAAGTGGCATGAGGGGACCGTGAAGCGGCCAGGTTTCGAGATCGAGGCGGAGG 163
Qy 99 CGAGGCGCGCCCTTACGAGGGGCCACACACCGTGAAGTTGAAGGTGACCAAGGCGCGCCC 158
Db 164 GAGTGGCAACCCAAAGACCGATCGTGAAGTAAACATGAAGTCTACCAAGGTG--C 220
Qy 159 CTTCCCTTCGCTGGGACATCTCTGCCCCGATTCACAGGCTCCAGGTTGACGT 218
Db 221 CTTCCCTTCGCTGGGACATCTCTGCCCCGATTCACAGGCTCCAGGTTGACGT 280
Qy 219 GAAGCACC CGCGCGACATCGCGACTACAAGAGCTCTCTCTCCCGAGGCGCTTCAAGTG 278
Db 281 ACCCTTCCTGACGCTCGCTGCGCTTTCAGGCGAGCTGTGAACACGCGGATCAAT 340
Qy 279 GGAGCGGCTGATGAATTCGAGGAGCGCGCGTGGCGACCGTGAACCGAGGACTCTCCCT 338
Db 341 GCATCGCTCTTCAACTTCGAGGACGCTGCGCTGCTGCTGAGGCGGACCTACTACTCTA 400
Qy 339 GCAGGAGCGCTGCTTCATCTACAGGTTGAGTTCATCGGCGTGAACCTCCCTCCGAGG 398
Db 401 CAGCGCGCGCAAGATCCATGAGAGTTTCATCTCGTGGGAGGCTGTTTCCCGGAGATAG 460
Qy 399 CCGCGTGTATGAGAGGAGGACCATGGGCTGGGA 431
Db 461 TCCGCTGATGACCAACGCGCTGACCGTTTGA 493

RESULT 4

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BI382442 707 bp mRNA linear EST 26-AUG-2003
BFLG2_000619 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498C08131 5',
mRNA sequence.

BI382442 1 GI:30917513

EST.

Branchiostoma floridae (Florida lancelet)

Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

Branchiostoma.

1 (bases 1 to 707)

Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,

Herwig, R., Vingron, M. and Lehrach, H.

New evidence for genome-wide duplications at the origin of

vertebrates using an amphioxus gene set and completed animal

genomes

Genome Res. 13 (6A), 1056-1066 (2003)

22683279

12799346

Contact: Panopoulou G

Laboratory 145, Dept. Lehrach

Max-Planck-Institut fuer Molekulare Genetik

Innestr. 63-73, D-14195 Berlin, Germany

Tel.: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: panopoul@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting

(ONFP) to reduce sequencing redundancy. According to the ONFP

procedure, clones giving the same hybridisation pattern with a

battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>. Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers

FORWARD: 5' CCCAGGCTTACACTTATGCTCCGGCTCG 3' (M13RSP)

BACKWARD: 5' GCTATTACGACCTGGCGAAGGGGAGTGTG 3' (M13FSP)

Insert Length: 1 Std Error: 0.00

Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86

High quality sequence stop: 707.

FEATURES

source

1..707

/organism="Branchiostoma floridae"

/mol_type="mRNA"

/db_xref="taxon:7739"

/clone="MPMGp498C08131"

/tissue_type="whole embryo"

/dev_stage="5-6 hrs (gastrula stage)"

/lab_host="E.coli, XLI blue"

/clone_lib="Amphioxus 5-6 hrs cDNA library (Name

convention: BFLG or MPMGP498)"

/note=vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');

Site 2: NotI, BamHI, HindIII (3'); OligodT primed and

directionally cloned in pSport1 vector using a NotI

(5'-pGACTAGTTCTAGATCGGAGCGCGCC (T)15-3' and a SalI 5'-

TCGACCCACCGCTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 11.4%; Score 77.6; DB 12; Length 707;

Best Local Similarity 49.3%; Pred. No. 0.11;

Matches 259; Conservative 0; Mismatches 260; Indels 6; Gaps 2;

```

QY 39 CTTCAGGTGGGATGAGGACCGGTGAACGCCACGAGTTCGAGATCGAGGCGGAGG 98
DB 133 CTTCAGGTGGGATGAGGACCGGTGAACGCCACGAGTTCGAGATCGAGGCGGAGG 192
QY 99 CGAGGGCGGCGCTACGAGGCGGACACACCGTGAAGTTGAAGGTGACCAAGGCGGCGCC 158
DB 193 CAGTGGCAACCCAAAGACGAGTTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 249
QY 159 CTGCGCCCTCGCTGGGACATCTGTCCTCCCGAGTTCGAGTGAAGTGAAGTGAAGTGAAGT 218
DB 250 CTGCGCGGTTTCTCTCTGTTGGGCGCGCATCTAGGTTAGGTTAGGTTAGGTTAGGTTAGG 309
QY 219 GAAGCACCGCGCGACATCCCGACATCAAGAGAGTGTCTTCCCGAGGCGTTCAGGTG 278
DB 310 ACCCTTCCTGATGGTTCGCTGCTTCGAGGATCCGTGAACACGCGGGTATCAAT 369
QY 279 GAGCGCGGTGATGAATTCGAGGACGCGCGGTGGGACCGTGAACGAGATCTCTCCCT 338
DB 370 GCATCGCTCTTCAACTTTGAGGACGCGTGCCTGACTGCCACCTCAACTACTCTCTA 429
QY 339 GCAGGAGCGGTGCTTCATCTACAAGGTGAAGTTCATCGGCGTGAAGTTCCTCCCTCGACGG 398
DB 430 CAGCGCGCGGCAAGATCAGGAGAGTTTCTCTGTTGGGACGAGTGTTCCTCCGACGATAG 489
QY 399 CCGCGTATCAGAAGAACACCTAGGCTGGG---GGCTCCACGAGCGGCTGTACCC 455
DB 490 TCCGGTATACCAACACGCGTACCGCTCTGGACAGGAGCGTGGCCAAAGTGTATGTGCT 549
QY 456 CCGGAGCGGCTGCTGAAGCGCGAGATCCACAGGCGCTGAAGTGAAGGACGCGGCA 515
DB 550 GTCCGATGACAAGTTTGGCCAGTTCGTGGACTGGACTTACCTCACCAGGCGGCGGGG 609
QY 516 CTACCTTGGTGGATTCAGTTCATCTACATGCGCCAAAGACCCGT 560
DB 610 CTACCGTGCCACGGTGCAGACCAACTTCACTTCGCANAGCCCAT 654

```

RESULT 5

B1379061

LOCUS

DEFINITION

B1379061 561 bp mRNA linear EST 26-AUG-2003
BFLG1_000692 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MPMGP498) Branchiostoma floridae cDNA clone MPMGP498P1084 5',
mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Genomes

Genome Res. 13 (6A), 1056-1066 (2003)

22683279

12799346

Contact: Panopoulou G

Laboratory 145, dept. Lehrach

Max-Planck-Institut fuer Molekulare Genetik

Innestr.63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: panopoul@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting

(ONFP) to reduce sequencing redundancy. According to the ONFP

procedure, clones giving the same hybridisation pattern with a

battery of 200 8mer oligonucleotides are grouped into clusters. One

clone per cluster is selected for sequencing. The size of each

cluster is an indicator of the frequency of a transcript in the

analysed library. The cluster size as well as the coordinates of the

rest of the clones assigned to the same fingerprint cluster as the

clone from which the above EST is generated is available at the

amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>

Clones and filters are distributed via the Resource Center/Primary

Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers

FORWARD: 5' CCCAGGCTTACACTTATGCTCCGGCTCG 3' (M13RSP)

BACKWARD: 5' GCTATTACGACCTGGCGAAGGGGAGTGTG 3' (M13FSP)

Insert Length: 1200 Std Error: 0.00

Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86

High quality sequence stop: 561.

Location/Qualifiers

1..561

/organism="Branchiostoma floridae"

/mol_type="mRNA"

/db_xref="taxon:7739"

/clone="MPMGp498P1084"

/tissue_type="whole embryo"

/dev_stage="5-6 hrs (gastrula stage)"

/lab_host="E.coli, XLI blue"

/clone_lib="Amphioxus 5-6 hrs cDNA library (Name

convention: BFLG or MPMGP498)"

/note=vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');

Site 2: NotI, BamHI, HindIII (3'); OligodT primed and

directionally cloned in pSport1 vector using a NotI

(5'-pGACTAGTTCTAGATCGGAGCGCGCC (T)15-3' and a SalI 5'-

TCGACCCACCGCTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 11.4%; Score 77.4; DB 12; Length 561;

Best Local Similarity 52.6%; Pred. No. 0.11;

Matches 193; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 42 CAAGGTGCGATGAGGCGCACCGGTGAACGCCACGAGTTCGAGATCGAGGCGGAGGCGGA 101

DB 92 CGACCTTCACATCTTCGGCTCCGTCACGCGCGGAGTTCGACTGCTGGGAGCGGAAA 151

COMMENT Contact: Panopoulou G
Laboratory 145, dept.Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACACTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GTATTACCCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert length: 1200 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86
High quality sequence stop: 671.
Location/Qualifiers
source
1. .671
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp498C047"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498)"
/note="vector: pSport1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTCTAGATCGGAGCGGCCGCC (T)15-3' and a SalI 5'-TCGACCCAGCGCGTCCG-3' adapters (Gibco BRL)."

Qy Query Match 11.4%; Score 77.4; DB 12; Length 671;
Best Local Similarity 52.6%; Pred. No. 0.12;
Matches 193; Conservative 0; Mismatches 171; Indels 3; Gaps 1;
42 CAAGGTCCGATGAGGCGACCGTGAACCGGCGGAGTTCGAGATCGAGGCGAGGGCGA 101
Db |||||
71 CGACCTTCACATCTTCGGCTCCGTCATATGCGCGGAGTTCGACTGTTGGAGGCGGAA 130
Qy |||||
102 GGGCGGCGCTTACGAGGCGGCACACCGGTGAAGTTGAAGTGAACCAAGGCGGCGCCCT 161
Db |||||
131 GGGCAACCCGACGATGGAACCGGTGGAGACCACTGTGAATCCACCGCGGCG-CCCT 187
Qy |||||
162 GCCTTCGCTGGACATCTGTCGCCCGGATTCAGTCCAGTCCAGTCCAGGTGTACGTGAA 221
Db |||||
188 GCGCTGCTCCCGCTGTGTGATCGGACCCCACTGGGGTACGTTTTCACGATCTGCGC 247
Qy |||||
222 GCACCCGCGGACATCCCGGACTACAGAAAGCTGTCTTCCCGAGGCGTTCAGTGGGA 281
Db |||||
248 CTTCCCTGGCGGCGCTCACTCCCTTCCAAACCGGCATCAGCGAGGAGGTACCGTTCA 307
Qy |||||
282 CGCGTGATGAATTCGAGAGCGGCGCGTGGGACCGGTGACCCAGATCTCTCCCTGCA 341
Db |||||
308 ACGTGTGTTCAAGTTTGAAGACCGCGGAGTGTGAATTGCACTTCGCTTACACCTACGA 367
Qy |||||
342 GGACGCGTGTTCATCTACAGAGTGAAGTTTCATCGGCGGTGAATCTCCCTCCGACGCGC 401
Db |||||
368 GGGGCGCAAGNTCAAAGGGAGTTCAGCTGATCGGTCAGGTTTCCCTCGCGCGGCGC 427
Qy |||||
402 CGTGATG 408
|||

Db 428 TGTGATG 434

RESULT 8
BI387917
LOCUS
DEFINITION
BI387917
687 bp mRNA linear EST 26-AUG-2003
BFL26.002620 Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531) Branchiostoma floridae cDNA clone MPMGP531N0767 5', mRNA sequence.
BI387917
ACCESSION
BI387917.1 GI:30922753
VERSION
KEYWORDS
SOURCE
ORGANISM
Branchiostoma floridae (Florida lancelet)
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.
1 (bases 1 to 687)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J., Herwig, R., Vingron, M. and Lehraach, H.
New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
Laboratory 145, dept.Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACACTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GTATTACCCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert length: 1200 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86
High quality sequence stop: 687.
Location/Qualifiers
source
1. .687
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531N0767"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531)"
/note="vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTCTAGATCGGAGCGGCCGCC (T)15-3' and a SalI 5'-TCGACCCAGCGTCCG-3' adapters (Gibco BRL)."

Qy - 42 CAAGGTCCGATGAGGCGGACCGGTGAACCGGCGGAGTTCGAGATCGAGGCGAGGGCGA 101

Db 95 CGACATCCACTCCAGGCTCATCAACGGCCACAGTTCGACATGTTGGGCGGAGGAAA 154
 QY 102 GGGCGGCCCCCTACGAGGCGCAACACCGTGAAGTTGAAGTGAACAAAGGCGGCCCCCT 161
 Db 155 AGGCGACCCGAAACGCGGCTCGCTGGTACACAGGAAATCCACCAAGGCT---CCGCT 211
 QY 162 GCCCTTCGCTGGGACATCCTCTCCCGAGTTCAGTACGCTCCAAAGTGTAGTGAA 221
 Db 212 GAAGTTCTCTCCCTACTTGTATATCCCGACCTCGGTCAGGTTACTACAGTACTCTCC 271
 QY 222 GCACCCCGCGACATCCCGACTACAAG---AAGTGTCTTCTCCCGAGGCTTCAAGTG 278
 Db 272 CTACCGGACGACCTCGCTTTCAGGCGCCACCATGTGGAAGAACAGGATATGCAGT 331
 QY 279 GAGGCGGTGATGAATTCGAGAGCGGCGCTGCGACCGTGAACCCAGGACTCTCTCCT 338
 Db 332 CCATCGCGTTCGACTTGAAGCGGAGGCAAGTGTCCATCGAGTTTAAGTACTCTCTA 391
 QY 339 GAGGACGCGCTGCTTCATCTACAAGGTGAAGTTTCATCGGCGTGAATTTCCCTCCGACGG 398
 Db 392 CGAGGGTTCCTATCAAGGCGGACATGAAGTTTACGGGACCGGTTTCCCTGAGGACGG 451
 QY 399 CCGGTGATGAGAGAACACATCGGCTGGGAGGCTTCC---ACGAGCGCTGTATCC 455
 Db 452 GCGGTGATGAGAGAACACATCGGCTGGGAGGCTTCC---ACGAGCGCTGTATCC 511
 QY 456 CGCGACGCGCTGCTGAAGGCGGAGATCCACAGCGCTGAGCTGAAGTGAAGGCGGCGCA 515
 Db 512 TACGACACACATCGTGGACAGCTTCGCTGGACTTACACCTTGCAGATGGGAAGCG 571
 QY 516 CTACCTGTGTGAGTTCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 562
 Db 572 CTCAAGGCGCAAGTGAAGGCGGAGTTCATCATCATCATCATCATCATCATCATCAT 618

RESULT 9

BI386952

LOCUS

DEFINITION

BI386952 459 bp mRNA linear EST 26-AUG-2003
 BFL26_001655 Amphioxus 26hr cDNA library (Name convention: BFL26 or
 MPMP531) Branchiostoma floridae cDNA clone MPMP531M0997 5', mRNA
 sequence.

ACCESSION

BI386952

VERSION

BI386952.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

clone from which the above EST is generated is available at the
 amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
 Clones and filters are distributed via the Resource Center/Primary
 Database of the German Genome Project (<http://www.rspd.de>)
 PCR Primers
 FORWARD: 5' CCCAGCTTTACACTTTATCTTCCGGCTCG 3' (M13RSP)
 BACKWARD: 5' GGTATTACCCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
 Insert length: 1 Std Error: 0.00
 Seq primer: 5'-CCGTCGCGAATTCGCGGT-3' pSport3/86
 High quality sequence stop: 459.

FEATURES

source

1. 459

Location/Qualifiers

/organism="Branchiostoma floridae"

/mol_type="mRNA"

/db_xref="taxon:7739"

/clone="MPMP531M0997"

/tissue_type="whole embryo"

/dev_stage="26 hrs (neutula stage)"

/lab_host="E.coli, XL1 blue"

/clone_lib="Amphioxus 26hr cDNA library (Name convention:
 BFL26 or MPMP531)"/notes="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI,
 EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); Oligo dT,
 primed and directionally cloned in pSport1 vector using a
 NotI (5'-PGACTAGTTCAGATCGAGCGGCCGCC (T)15-3' and a
 SalI 5'-TCGACCCAGCGCTCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 11.3%; Score 76.4; DB 12; Length 459;
 Best Local Similarity 52.3%; Pred. No. 0.15;
 Matches 192; Conservative 0; Mismatches 172; Indels 3; Gaps 1;
 QY 42 CAAGTCCGATGAGGCGACCGTGAACGCGCCACGAGTTCGAGATCGAGGCGGAGGCGGA 101
 Db 79 CGACCTTCACATCTTCGGCTCCGTCATGCGCGGAGTTCGACTGTGTGGAGGCGGAAA 138
 QY 102 GGGCGGCGCTAGAGGGGCCACACACCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 161
 Db 139 GGGCAACCCGACGATGGAACCGTGGAGACAGTGTGAATCCACCGGGGCG---CCCT 195
 QY 162 GCCCTTCGCTGGGACATCTCTGCCCCCAGTTCACAGTACGGCTCCAAAGTGTACGTGAA 221
 Db 196 GCCCTGCTCCCGCTGCTGATCGAGCCCAACCTGGGTACGGCTTCTACCACTACCTGCC 255
 QY 222 GCACCCGCGGACATCCCGGCTACAGAGAGTGTCTCTCCCGAGGCTTCAAGTGGGA 281
 Db 256 CTTCCCTGGGGCGGCTCGCCCTTCCAAACAGCCATCACGAGCGAGGTACCAAGTTC 315
 QY 282 GCGGTGATGAATTCGAGGACGCGGCGGTGGCGACCGTGAACCCAGGACTCCTCCCTGCA 341
 Db 316 ACGTGTTCAGTTTGAAGACCGGGAGTGTGAATTGCAACTTCGCTACACTACNA 375
 QY 342 GGACGGTGTTCATCTPACAAGTGAAGTTCATCGGCGTGAATCTCCCTCCGACGCGCC 401
 Db 376 GGGCGGTGAAGATCAAGGGGAGTTCAGCTGATCGGTGAGTTCCTCCCGCGGCGGCC 435
 QY 402 CGTGATG 408
 Db 436 TGTGATG 442

RESULT 10

BI385436

LOCUS

DEFINITION

BI385436 561 bp mRNA linear EST 26-AUG-2003
 BFL26_000139 Amphioxus 26hr cDNA library (Name convention: BFL26 or
 MPMP531) Branchiostoma floridae cDNA clone MPMP531H0446 5', mRNA
 sequence.

ACCESSION

BI385436

VERSION

BI385436.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BI385436 561 bp mRNA linear EST 26-AUG-2003
 BFL26_000139 Amphioxus 26hr cDNA library (Name convention: BFL26 or
 MPMP531) Branchiostoma floridae cDNA clone MPMP531H0446 5', mRNA
 sequence.
 ACCESSION
 BI385436
 VERSION
 BI385436.1
 GI:30920419
 EST.
 Branchiostoma floridae (Florida lancelet)
 Branchiostoma floridae
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

Email: panopoul@molgen.mpg.de
 The library was characterised by oligonucleotide fingerprinting
 (ONFP) to reduce sequencing redundancy. According to the ONFP
 procedure, clones giving the same hybridisation pattern with a
 battery of 200 8mer oligonucleotides are grouped into clusters. One
 clone per cluster is selected for sequencing. The size of each
 cluster is an indicator of the frequency of a transcript in the
 analysed library. The cluster size as well the coordinates of the
 rest of the clones assigned to the same fingerprint cluster as the

Branchiostoma.
1 (bases 1 to 661)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
Herwig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACACTTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GGTATACCGCAGCTGGGAAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86
High quality sequence stop: 661.
Location/Qualifiers
1. 661
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531H0446"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neural stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention:
BFL26 Or MPMGP531)"
/notes="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI,
EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT,
primed and directionally cloned in pSport1 vector using a
NotI (5')-pGATCTTCTAGATCGGAGCGGCCCC (T)15-3' and a
SalI 5'-TCGACCCACGCTCGC-3' adapters (Gibco BRL)."
ORIGIN
Query Match 11.2%; Score 75.8; DB 12; Length 661;
Best Local Similarity 50.5%; Pred. No. 0.2;
Matches 266; Conservative 0; Mismatches 252; Indels 9; Gaps 3;
QY 42 CAAGGTGCGATGAGGGCAACCGTGAACGCCACGAGTTCGAGATCGAGGGCGAGGGA 101
Db 61 CGACATCCACCTCCAGGCTCCATCAACGCCACGAGTTCGACCTGGTGGGGGAGGACA 120
QY 102 GGCGCCCTCTAGAGGGCCACAACACCGTGAAGTTCAGGTGACCAAGGGCGGCCCT 161
Db 121 AGCGACCGAAGCCCGGCTCGCTGGTGACCAAGCGAATCACCACAGGGT---CCCT 177
QY 162 GCCTTTCGCTGGGACATCTGTCCCGCCAGTTTCCAGTACGGTCCAAAGGTGTAGTGAA 221
Db 178 GAAGTTCTCTCCCTACTTGTATGATCCCCACCTCGGGTACGGGTACTACAGTACCTCCC 237
QY 222 GCACCCCGGACATCCCCGATACAG---AGCTGTCTTCCCGAGGGCTTCAAGTG 278
Db 238 CTACCCGGAGGACCCCTCGCTTTCAGGCGCCACCATGTTGGGAAGGATCAGGGTATGACGT 297

QY 279 GGAGCGGTGATGAACCTTCGAGGACGGCGGTGGACCGTGCACCGAGCTCCTCCCT 338
Db 298 CCATCGGTGTTCGACTTTGAAGACGGGACGAGCTGTCCATGAGTTTAAATACCTTA 357
QY 339 GCAGGACGGTGTCTTCATCTACAAGGTGAAGTTTCATCGCGGTGAACCTTCCCTCCGACGG 398
Db 358 CGAGGTTCCATATCAAGCCGACATGAAGTTTCAAGGAAACCGGTTTCCCTGAGGACGG 417
QY 399 CCCGCTGATGCAAGAAGACCATGGCTGGAGGCTCC---ACCGAGCGCTGTACCC 455
Db 418 GCGGTGATGACCAAGCCAGATTGTCGACCAAGGACGGCTGCTGTCCAAAGAACCTTACCT 477
QY 456 CCGGACGGGTGCTGAAGGGCGAGATCCCAAGGCCCTGAAGCTGAAGGACGCGCGCCA 515
Db 478 TAACGACAAACACCATCGTGGACAGCTTCGACTGACTTACACCTGCAGATGGAGCG 537
QY 516 CTACTGTTGGAGTTCAGTCCATCTATCATGGCCCAAGACCGCGTGC 562
Db 538 CTACAAGGCCCAAGTGCAGGACCACTACATCTTCGGCAAGCCCTTTC 584

RESULT 11

LOCUS B1379468
DEFINITION B1379468 637 bp mRNA linear EST 26-AUG-2003
or MPMGP498) Branchiostoma floridae cDNA clone MPMGP49801918 5',
mRNA sequence.
ACCESSION B1379468
VERSION B1379468.1 GI:30914617
SOURCE EST.
ORGANISM Branchiostoma floridae (Florida lancelet)
KEYWORDS Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.
1 (bases 1 to 637)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
Herwig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACACTTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GGTATACCGCAGCTGGGAAAGGGGATGTG 3' (M13FSP)
Insert Length: 1200 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86
High quality sequence stop: 637.
Location/Qualifiers
1. 637
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"

FEATURES
source

/clone="MPMGp49801918"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498)"
/note="Vector: pSPort1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSPort1 vector using a NotI (5')-pGACTAGTTCTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-TCGACCCACGGGTCGG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 11.1%; Score 75.4; DB 12; Length 637;
Best Local Similarity 50.5%; Pred. No. 0.22;
Matches 265; Conservative 0; Mismatches 251; Indels 9; Gaps 3;
QY 42 CAAGTGGCGCATGGAGGCGACCGTCAACGGCCACGAGTTTCGAGATTCGAGGCGGAGGGCGGA 101
Db 78 CGAGATCCACCTCCACGTTCCATCAACGGCCACGAGTTTCGAGATTCGAGGCGGAGGGCGGA 137
QY 102 GGGCGCGCTTCAGAGGGCCACACCGTGAAGTTGAAGTGACCAAGGGGGGGGGGGGGGGGGGG 161
Db 138 AGGCGACCCAAACGGCGGCTCGCTGTGACCAACGAGAAATCCACCCAG---GGTCCCGCT 194
QY 162 GCCTTTCGCTGGGACATCTGCTCCCGCAGTTCAGTACGCTCCAGGTCCTCAAGGTGTACGTGAA 221
Db 195 GAAGTCTCTCCGCTTGTGATCCCGCAGTTCAGTACGCTCCAGGTCCTCAAGGTGTACGTGAA 254
QY 222 GCACCCCGCCGACATCCCGACTCAAG---AAGTGTCTTTCCTCCCGAGGCGGTTCAGGTG 278
Db 255 CTACCCGGACGGACCTCGCTTTCAGGCGCACCATGTTGGAAGGATCGGGGTATACAGT 314
QY 279 GAGCGCGTGAAGTTCGAGCGGGCGGCGTGGCGACCGTTCAGGTCCTCCCGAGGCGGTTCAGGTG 338
Db 315 CCACCGCGTGTGTGATTCGAAGCGGAGGCGAGCTGTCATCGAGTTCAGTACTCTCTA 374
QY 339 GAGGACCGCGTCTTCATCTAAGGTGAAGTTTCATCGGCGTGAAGTTCCTCCCGAGGCGG 398
Db 375 CGAGGGTTCCTATTAAGCGCGACATGAAGTTTCAGGGAACCGGTTTCCCTGAGGATGG 434
QY 399 CCGCGTGTGAGAGAACCATCGGCTGGAGGCGCTCC---ACGAGCGCTCTAGCC 455
Db 435 CGCGGTTCATGACCGACGAGTGTTCGACGAGGCGGCTCGTGTTCAGAGAACACCTACTCT 494
QY 456 CGCGGACCGCGTCTGAAGGGCGAGATCCACGAGCGCTGAAGCTGAAGGAGCGGCGGCA 515
Db 495 CACGACACACCATCTGTCGACCACTTCGACTGACTTACACTTCGAGATTCGAGATTCGAGG 554
QY 516 CTACCTGTGTGAGTTCAAGTCCATCTACATGCGCCAGAGAGCGCGT 560
Db 555 CTACAGGGCGGAGTGACGAGGCGACTTACATCTTCGACAGAGCGCTT 599

RESULT 12

BI382373
LOCUS BFLG2 000550 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498) Branchiostoma floridae cDNA clone MPMGP498C1177 5', mRNA sequence.
DEFINITION
ACCESSION BI382373
VERSION BI382373.1 GI:30917444
KEYWORDS EST.
SOURCE Branchiostoma floridae (Florida lancelet)
ORGANISM Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.
REFERENCE 1 (bases 1 to 591)
AUTHORS Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A., J., Herwig, R., Vingron, M. and Lehrach, H.
TITLE New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes

JOURNAL
MEDLINE
PUBMED
COMMENT

Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>.
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACATTTATGCTCCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACCCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGTCGGAATTCGCGGT-3' pSPort3/86
High quality sequence stop: 591.

FEATURES

source

1..591
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp498C1177"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498)"
/note="Vector: pSPort1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSPort1 vector using a NotI (5')-pGACTAGTTCTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-TCGACCCACGGTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 10.7%; Score 72.6; DB 12; Length 591;
Best Local Similarity 51.8%; Pred. No. 0.55;
Matches 190; Conservative 0; Mismatches 174; Indels 3; Gaps 1;
QY 42 CAAGTGGCGATGGAGGCGACCGTGAACGGCCACGAGTTCGAGATCGAGGCGGAGGGCGGA 101
Db 91 CGACCTTCACATCTTCGGCTCCGTCATGCGGAGGATTCGACCTGTGTGGAGCGGAAA 150
QY 102 GGGCGCGCCCTACGAGGCGCACACCGTGAAGTTGAAGTTCACCAAGGCGGCGCCCT 161
Db 151 GGGCAACCGGATGATGAGACCGTGGAGACCATGTCGAATCCACCGGGGCG--CGCT 207
QY 162 GCCTTCGCTGGGACATCTGTCCTCCCGAGTTCAGTCCAGTCCAGGCTCCAGGTTGACGTGAA 221
Db 208 GCCTTCGCTCCCGCTGTTGATCGGACCCCAACTGGGGTACGGCTTCTACCACTACCTGCC 267
QY 222 GCACCGCGCGACATCCCGACTACAAGAGCTGTCTCTCCCGAGGCTTCAAGTGGGA 281
Db 268 CTTCCCTGGCGCGCTTACCCCTTCCAAACCGCCATCAGGACGAGGGTACAGGTTC 327
QY 282 CGCGGTGATGAACCTTCGAGGACCGCGGCGTGGGACCGGTGACCCAGGACTCTCCCTGCA 341
Db 328 CCGTGTGTTCAAGTTTGAAGACGGAGGATTGCTGAATTGCAATTCGCGTACACCTACGA 387
QY 342 GAGCGGTGCTTCATCTACAGGTGAGTTCATCGGCGTGAACCTTCCCTCCGACGGCC 401
Db 388 GGGCGGCAAGATCAAAAGGGGAGTTCACGCTGATCGGTGAGTTTCTCTCCGCGCGGCC 447

QY 402 CQTGATG 408
 Db 448 TGTGATG 454

RESULT 13
 BI382638 675 bp mRNA linear EST 26-AUG-2003
 LOCUS BFLG2.000815 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
 or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498F0258 5',
 mRNA sequence.
 ACCESSION BI382638
 VERSION BI382638.1 GI:30917704
 KEYWORDS EST.
 SOURCE Branchiostoma floridae (Florida lancelet)
 ORGANISM Branchiostoma floridae
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 Branchiostoma.
 1 (bases 1 to 675)
 Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
 Herwig, R., Vingron, M. and Lehrach, H.
 New evidence for genome-wide duplications at the origin of
 vertebrates using an amphioxus gene set and completed animal
 genomes
 Genome Res. 13 (6A), 1056-1066 (2003)
 JOURNAL 22683279
 MEDLINE 12799346
 PUBMED
 COMMENT Contact: Panopoulou G
 Laboratory 145, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Innestr. 63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1235
 Fax: +49 30 8413 1128
 Email: panopoul@molgen.mpg.de
 The library was characterised by oligonucleotide fingerprinting
 (ONFP) to reduce sequencing redundancy. According to the ONFP
 procedure, clones giving the same hybridisation pattern with a
 battery of 200 8mer oligonucleotides are grouped into clusters. One
 clone per cluster is selected for sequencing. The size of each
 cluster is an indicator of the frequency of a transcript in the
 analysed library. The cluster size as well the coordinates of the
 rest of the clones assigned to the same fingerprint cluster as the
 clone from which the above EST is generated is available at the
 amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
 Clones and filters are distributed via the Resource Center/Primary
 Database of the German Genome Project (<http://www.rzpd.de>)
 PCR Primers
 FORWARD: 5' CCCAGCCTTACATTTGCTCCGCTCG 3' (M13RSP)
 BACKWARD: 5' GTATTACCCAGCTGGGAAGGGGATGTG 3' (M13FSP)
 Insert length: 1 Std Error: 0.00
 Seq primer: 5'-CCGTCGGGAATCCCGGGT-3' pSport3/86
 High quality sequence stop: 675.
 Location/Qualifiers
 1..675
 /organism="Branchiostoma floridae"
 /mol_type="mRNA"
 /db_xref="taxon:7739"
 /clone="MPMPGp498F0258"
 /issue_type="whole embryo"
 /dev_stage="5-6 hrs (gastrula stage)"
 /lab_host="E.coli, XLI blue"
 /clone_lib="amphioxus 5-6 hrs cDNA library (Name
 convention: BFLG or MPMPGp498)"
 /notes="vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
 Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
 directionally cloned in pSport1 vector using a NotI
 (5'-pGACTAGTCTAGATCGGCGCGGCC (T)15-3' and a SalI 5'-
 TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

FEATURES
 source
 1..675
 /organism="Branchiostoma floridae"
 /mol_type="mRNA"
 /db_xref="taxon:7739"
 /clone="MPMPGp498F0258"
 /issue_type="whole embryo"
 /dev_stage="5-6 hrs (gastrula stage)"
 /lab_host="E.coli, XLI blue"
 /clone_lib="amphioxus 5-6 hrs cDNA library (Name
 convention: BFLG or MPMPGp498)"
 /notes="vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
 Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
 directionally cloned in pSport1 vector using a NotI
 (5'-pGACTAGTCTAGATCGGCGCGGCC (T)15-3' and a SalI 5'-
 TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

ORIGIN
 Query Match 10.3%; Score 69.6; DB 12; Length 675;
 Best Local Similarity 49.7%; Pred. No. 1.5;

Matches 261; Conservative 0; Mismatches 255; Indels 9; Gaps 3;
 QY 42 CAAGTGCATGGAGGCCACCGTGAACGGCCACAGTTCGAGATCGAGGCGAGGCGA 101
 Db 106 CGACATCCACCTTCACGGCTCCATCAACGGCCACAGTTCGACATGCTGGGGGAGGAAA 165
 QY 102 GGGCGCCCTACGAGGCCACACCGTGAAGTTGAAGTGACCAAGGCGGCGCCCT 161
 Db 166 AGGCGACCCGACCGCGCTCGCTGGTGACACACAGCGAAATCCACCAAGGT---CCCT 222
 QY 162 GCCCTTCGCTGGGACATCTCTGTCCTCCCGGCTTCAGTACGGCTCCCAAGGTGTACGTGAA 221
 Db 223 GAAGTTCTCCCTACTTGTATGATCCCTCCCGGCTACCTACAGTACCTCCC 282
 QY 222 GCACCCCGGACATCCCGGCTACAGAGCTGCTCTTCCCCGA---GGCTCAAGTG 278
 Db 283 CTACCCGACGACCTCGCTTTCAGGTTCCTCATGTTGGAAGGATCGGGGTATGCAGT 342
 QY 279 GGAGCGGTGATGAATTCGAGGACGGCGCTGGGACCGTGACCCAGGACTCCTCCCT 338
 Db 343 CTACCGGTGTTCCACTTTGAGACGGGAGCAAGCTGACTACCGAGTTTAAGTACTCTA 402
 QY 339 GCAGACGGCTGCTTCATCTACAAGGTGAAGTTTCATCGGGGTGAATTTCCCTCCCGACGG 398
 Db 403 CGAGGGTTCCCATATCAAGGCCGACATGAAGCTGATGGGAAGCGGTTTCCCTGACGACGG 462
 QY 399 CCCGTGATGCAGAGAAGACCATGGGCTGGAGGCTCC---ACCGAGCGCTGTACCC 455
 Db 463 CCCAGTATGACGACGACGATTTGCGACGAGCGGCTGCTGTCGAAGAGGAGTATCT 522
 QY 456 CCGGACGCGGTGCTGAAGGCGGAGATCCACAAGGCCCTGAAGCTGAAGACGCGCGCCA 515
 Db 523 TACAACAACACCATCGTGACAGCTTCGACTGGAGTTACAACCTGCAGAAATGGGAAGCG 582
 QY 516 CTACTCGTGGAGTTCAAGTCCATCTACATGGCCAGAGCCCGT 560
 Db 583 CTACAGGCGCGGTGTCNAGCCACTACTCTTCGACAAGCCCTT 627

RESULT 14
 LOCUS BI376909
 DEFINITION BFLG3.000712 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
 or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498C2412 5',
 mRNA sequence.
 ACCESSION BI376909
 VERSION BI376909.1 GI:30911858
 KEYWORDS EST.
 SOURCE Branchiostoma floridae (Florida lancelet)
 ORGANISM Branchiostoma floridae
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 Branchiostoma.
 1 (bases 1 to 537)
 Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
 Herwig, R., Vingron, M. and Lehrach, H.
 New evidence for genome-wide duplications at the origin of
 vertebrates using an amphioxus gene set and completed animal
 genomes
 Genome Res. 13 (6A), 1056-1066 (2003)
 JOURNAL 22683279
 MEDLINE 12799346
 PUBMED
 COMMENT Contact: Panopoulou G
 Laboratory 145, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Innestr. 63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1235
 Fax: +49 30 8413 1128
 Email: panopoul@molgen.mpg.de
 The library was characterised by oligonucleotide fingerprinting
 (ONFP) to reduce sequencing redundancy. According to the ONFP
 procedure, clones giving the same hybridisation pattern with a
 battery of 200 8mer oligonucleotides are grouped into clusters. One
 clone per cluster is selected for sequencing. The size of each
 cluster is an indicator of the frequency of a transcript in the
 analysed library. The cluster size as well the coordinates of the
 rest of the clones assigned to the same fingerprint cluster as the
 clone from which the above EST is generated is available at the
 amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
 Clones and filters are distributed via the Resource Center/Primary
 Database of the German Genome Project (<http://www.rzpd.de>)
 PCR Primers
 FORWARD: 5' CCCAGCCTTACATTTGCTCCGCTCG 3' (M13RSP)
 BACKWARD: 5' GTATTACCCAGCTGGGAAGGGGATGTG 3' (M13FSP)
 Insert length: 1 Std Error: 0.00
 Seq primer: 5'-CCGTCGGGAATCCCGGGT-3' pSport3/86
 High quality sequence stop: 675.
 Location/Qualifiers
 1..675
 /organism="Branchiostoma floridae"
 /mol_type="mRNA"
 /db_xref="taxon:7739"
 /clone="MPMPGp498F0258"
 /issue_type="whole embryo"
 /dev_stage="5-6 hrs (gastrula stage)"
 /lab_host="E.coli, XLI blue"
 /clone_lib="amphioxus 5-6 hrs cDNA library (Name
 convention: BFLG or MPMPGp498)"
 /notes="vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
 Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
 directionally cloned in pSport1 vector using a NotI
 (5'-pGACTAGTCTAGATCGGCGCGGCC (T)15-3' and a SalI 5'-
 TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>. Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers
FORWARD: 5' CCCCAGGTTTACACTTATGTTCCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTAGCCAGCTTGGGAAGGGGATGTG 3' (M13FSP)
Insert length: 1200 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86
High quality sequence stop: 537.

Location/Qualifiers
1. 537
/organism="Branchiostoma floridae"
/mol_type="rRNA"
/db_xref="taxon:7739"
/clone="MPMGp498C24112"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or WPMGp498)"
/note="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-GGACTAGTTCTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-TCGACCCAGCGCTCGG-3' adapters (Gibco BRL)."

ORIGIN		Query Match	10.2%; Score 69.4; DB 12; Length 537;
		Best Local Similarity	51.2%; Pred. No. 1.6;
		Matches	188; Conservative 0; Mismatches 176; Indels 3; Gaps 1;
QY	42	CAAGGTGCGCATGGAGGCGACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGA	101
Db	74	CGACCTTCAATCTTCGGCTCCGTCATGTGTCGGAGTTCGACCTGTTGGAGAGGAGAAC	133
QY	102	GGGCGCCCGCTTACGAGGCGCAACAACCGTGAAGTTGAAGGTGACCAAGGGCGGCCCT	161
Db	134	GGGCAACCCGACCATGGAACCGTGGAGACCAGTGTGAATCCACCCGGGGGG---CCCT	190
QY	162	GCCTTCGCTGGGATCCTGTCCCGCCAGTTCAGTACGGCTCCAGGTGTACGTGAA	221
Db	191	GTCTCTGCTCCCGCTGCTGATCGGACCCAACTTGGGTACGGCTTACAGTACTGCGC	250
QY	222	GCACCCCGCGCATCCCGACTACAAGAGTGTCTCTTCCCGAGGGCTTCAAGTGGGA	281
Db	251	CTTCCCTGGGGCGCTCGCCCTTCCAAACCGCCATCATTCAGCGAGGTTACAGGTTCA	310
QY	282	GCGCGTGATGAATTCGAGGACGGCGGTGGCGACCGTGACCCAGGACTCTCCCTGCA	341
Db	311	CCGTGTGTTCAGTTTGAAGACGGAGGAATGCTGGAATTCGAATTCCTGCTACACCTACGA	370
QY	342	GGACGGCTGCTTCACTACAAGGTGAAGTTTCATCGGGGTGAATTCCTCCCGACGGCCC	401
Db	371	GGCGGCGAAGATCAAGGGAGTTCACCTGTCTGGGTACAGGTTTCCTTCGCGGGGGCC	430
QY	402	CGTGATG 408	
Db	431	TG-TGATG 437	

RESULT 15
CG448047
LOCUS
DEFINITION
OGVGS13TH ZM 0.7 1.5 KB Zea mays genomic clone ZM8Ma0478D01,
genomic survey sequence.
ACCESSION
CG448047
VERSION
CG448047.1 GI:34833047
KEYWORDS
GSS.
SOURCE
Zea mays

ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
1. (bases 1 to 826)
Whitlaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nurnberg, A., Robbins, D. and Lakey, N.

TITLE
JOURNAL
COMMENT
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGVGS13TV
Contact: Cathy Whitelaw

FEATURES
Location/Qualifiers
1. 826
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM8Ma0478D01"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pECSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN		Query Match	10.2%; Score 69.4; DB 29; Length 826;
		Best Local Similarity	50.4%; Pred. No. 1.7;
		Matches	203; Conservative 0; Mismatches 191; Indels 9; Gaps 1;
QY	259	TTCCCCAGGGCTTCAAGTGGAGCGCGTGAATGAACCTTCAGAGCGCGGGTGGCGACC	318
Db	170	TTCCGCAAGGTGTTCAAGGGGGTCTGTCTCCAAAGTCCCGCACGGAGGTCCCGTGAACGG	229
QY	319	GTGACCCAGGACTCTCCCTGCAGAGCGGTCTTCACTACAGGTGAAGTTTCATCGGC	378
Db	230	GTGTCACAGACTCCACGAGGGCTCAGGAGTTCATCTCCGAGGTCTGTCAGATCGGC	289
QY	379	GTGAACCTTCCCTCCGACGGCCCGTGTATGCGAAGAACCATGCGGTGGAGGCGCTCC	438
Db	290	CACCTCCGCCACCGCAACCTCGTCAGCTGTGGGTACTCGCCGAGGAAGG-----	342
QY	439	ACCGAGCGCTGTACCCCGGACGGCGTGTGAGGGCGAGATCCACAGGCCCTGAAG	498
Db	343	--CGAGCTGTCTGTGTTACGACTACATGCCAAACGGCAGCTCGACAGCACTGCAC	400
QY	499	CTGAAGGACGGCGCCACTTACTGTGAGTTCAGTCCATCTACATGGCCCAAGAGCCC	558
Db	401	GGCGAGGACGCAAGCCGCTCTGAGTGGCGCAGCGGTTCAGATCGTCAAGGACGTG	460
QY	559	GTGACGCTGCCCGGCTACTACTAGTGGACACCAAGCTGGACATCACTCCCAACAGAG	618
Db	461	GGTCCGCGCTCTTCTACTCTCCAGAGAGTGGGAGGAGTGTGTCGTCCACCGGACGTC	520
QY	619	GACTACACCATCTGGAGCAGTACGAGCGCACCGAGGGCGGCC 661	
Db	521	AAGCCAGCAACGTCTCTCTCGACGGCGCATGTGTGCCCCACC 563	

Search completed: August 3, 2004, 15:47:59
Job time: 3043.03 secs

Blank Sheet US 'PTD

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:23:50 ; Search time 48.2188 Seconds
(without alignments)
7945.752 Million cell updates/sec

Title: US-10-081-864-14

Perfect score: 1237

Sequence: 1 atggcctctctcgagacgt.....gccaccaccttctctgttaa 678

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp

-Q=/cgn2_1/USPTO.spool_P/US10081864/runat_29072004.150505.25453/app_query.fasta_1.1742

-DB=A.Geneseq 29Jan04 -OFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEXT=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi

-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10081864 -CGN 1.1.128 -runat_29072004.150505.25453 -NCPU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSP=LOCK-100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.Geneseq 29Jan04:*

1: Genesep1980s:*

2: Genesep1980s:*

3: Genesep2000s:*

4: Genesep2001s:*

5: Genesep2002s:*

6: Genesep2003as:*

7: Genesep2003bs:*

8: Genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1214	98.1	225	5	Aae28837 Discosoma
2	1214	98.1	225	5	Aae17542 Discosoma
3	1204	97.3	225	5	Aae28836 Discosoma
4	1199	96.9	225	5	Aae28921 Discosoma
5	1199	96.9	225	5	Aae17541 Discosoma
6	1196	96.7	225	5	Aae28925 Discosoma
7	1195	96.6	225	5	Aae28919 Discosoma
8	1195	96.6	225	5	Aae28920 Discosoma
9	1195	96.6	225	7	Abw00937 Discosoma
10	1191	96.3	225	3	Aay99836 Discosoma

RESULT 1

AAE28837

ID AAE28837 standard; protein; 225 AA.

XX AC AAE28837;

XX DT 27-DEC-2002 (first entry)

XX DE Discosoma sp. drFP583 (NFP-6) mutant protein, E5-NA.

XX KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;

XX KW fluorescence activated cell sorting application; fluorescent timer;

XX KW biosensor; fluorescence resonance energy transfer application; FRET;

XX KW colouring agent; recombinant DNA application; analyte detection assay;

XX KW sunsreen; second messenger detector; drFP583 protein; NFP-6; muten;

XX OS Discosoma sp.

XX OS Synthetic.

XX FH Key

XX FH Location/Qualifiers

FT FT Misc-difference 2 /note= "Wild type Arg substituted with Ala"

FT FT Misc-difference 5 /note= "Wild type Lys substituted with Glu"

FT FT Misc-difference 9 /note= "Wild type Lys substituted with Thr"

FT FT Misc-difference 105 /note= "Wild type Val substituted with Ala"

FT FT Misc-difference 197 /note= "Wild type Ser substituted with Thr"

XX WO200268459-A2.

ALIGNMENTS

11	1191	96.3	225	3	AAB01622	Discosoma
12	1191	96.3	225	4	AAG5509	Anthozoon
13	1191	96.3	225	5	ABO8834	Yeast opt
14	1191	96.3	225	5	AAE28926	Discosoma
15	1191	96.3	225	5	AAE28833	Discosoma
16	1191	96.3	225	5	AAE17540	Discosoma
17	1191	96.3	225	5	AAO18270	Discosoma
18	1191	96.3	225	6	AAE34962	Discosoma
19	1191	96.3	225	7	ADC24126	Discosoma
20	1191	96.3	225	7	ABW00918	Discosoma
21	1191	96.3	487	5	ABO8821	Autofluor
22	1191	96.3	506	5	ABO8822	Autofluor
23	1191	96.3	547	5	ABO8823	Autofluor
24	1188	96.0	225	5	AAE28924	Discosoma
25	1188	96.0	225	7	ABW00929	Discosoma
26	1188	96.0	225	7	ABW00938	Discosoma
27	1187	96.0	225	7	ABW00930	Discosoma
28	1187	96.0	226	4	AAG5510	Anthozoon
29	1187	96.0	226	5	ABO8835	Yeast opt
30	1187	96.0	242	7	ADE24109	Discosoma
31	1186	95.9	225	5	AAE28922	Discosoma
32	1186	95.9	225	7	ABW00936	Discosoma
33	1186	95.9	225	7	ABW00931	Discosoma
34	1185	95.8	225	7	ABW00932	Discosoma
35	1185	95.8	225	7	ABW00935	Discosoma
36	1185	95.8	225	7	ABW00939	Discosoma
37	1184	95.7	225	7	ABW00941	Discosoma
38	1184	95.7	225	7	ABW00940	Discosoma
39	1183	95.6	225	7	ABW00933	Discosoma
40	1183	95.6	225	7	ABW00934	Discosoma
41	1182	95.6	240	6	ABP56678	Mammalian
42	1181	95.5	225	5	AAE28923	Discosoma
43	1179	95.3	240	6	ABP56685	Discosoma
44	1177	95.1	225	5	ABP70039	Colour Fa
45	1176	95.1	240	6	ABP56684	Discosoma

US 10 001 004 11:19

PD 06-SEP-2002.
 XX 20-FEB-2002; 2002WO-US005749.
 XX 21-FEB-2001; 2001US-0270983P.
 PR 04-DEC-2001; 2001US-00006922.
 XX (CLON-) CLONTECH LAB INC.
 PA Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Pradkov A;
 PI WPI; 2002-691654/74.
 XX N-PSDB; AAD46282.
 DR New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 PT of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
 PT analyte detection assays or fluorescence activated cell sorting
 PT applications.
 XX
 XX Disclosure; Page: 80pp; English.
 PS The invention relates to nucleic acid molecules encoding non-aggregating
 CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
 CC invention. Note: This sequence is not shown in the specification, but is
 CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
 CC ID NO:8 (AAE28833) in page 70-71 of the specification
 XX
 XX Sequence 225 AA;
 SQ

Alignment Scores:
 Pred. No.: 4.99e-102 Length: 225
 Score: 1214.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.14% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x AAE28837 (1-225)

QY 1 ATGGCTCTCCGAGAACGTCATCCGAGCTTCATCGCTTCAAGGTGGCGCATGGAGGC 60
 DB 1 MetAlaSerSerGluAsnValIleThrGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACCGCCACGAGTTCGAGATCGAGCGGCGAGCGGCGCCCTACGAGGC 120
 DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACAACACCGTGAAGTTGAAGGTGACACAGCGCGCGCCCTCGCTTCCTGGGACATC 180
 DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCCCCCAGTTCAGTACGCTCCAGGTGAGTGTAGTGAAGCACCCCGCGACATPCCC 240
 DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTACAGAAGTGTCTTCCCGAGGCGTTCAGTGGGAGCGCGTGTGATGAATTCGAG 300
 DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGCGCGTGGCGACCGGTGACCCAGGACTCTCTCCCTCAGAGCGGCTCTCATCTAC 360
 DB 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGGTGAAGTTCATCGCGTGAATTCCTCCGAGCGCGCCCGTGTGATGAGAGAGACC 420
 DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

QY 421 ATGGGTGGAGGCTCCACCGAGCGCTGTACCCCGGAGCGGCTGCTGAAGGGCGAG 480
 DB 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATCCACAGGCGCTGAAGCTGAAGGACGCGCGCCACTACCTGGTGGAGTTCAAGTCCATC 540
 DB 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TCATGGCCAAAGACCGCTGCGAGTCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
 DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspThrLysLeuAsp 200
 QY 601 ATCACCCTCCCAACGAGGACTACACCATCTGGAGGAGTACGAGCGCAGCGGCGGC 660
 DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACCACCTGTTCCTG 675
 DB 221 HisHisLeuPheLeu 225

RESULT 2
 AAE17542
 ID AAE17542 standard; protein; 225 AA.
 XX AAE17542;
 XX 22-APR-2002 (first entry)
 DT Discosoma sp. non-aggregating mutant timer protein ESNA.
 XX Fluorescent timer protein; protein movement; translocation; trafficking;
 KW promoter activity; gene expression; transgenic plant; gene modification;
 KW protein age; mutant; ESNA; mutein.
 XX Discosoma sp.
 OS Synthetic.
 OS Key Location/Qualifiers
 FH Misc-difference 2 /note= "Wild type Arg substituted with Ala"
 FT Misc-difference 5 /note= "Wild type Lys substituted with Glu"
 FT Misc-difference 9 /note= "Wild type Lys substituted with Thr"
 FT Misc-difference 9 /note= "Wild type Lys substituted with Thr"
 XX WO200196373-A2.
 PN 20-DEC-2001.
 PD 13-JUN-2001; 2001WO-US019097.
 PF 14-JUN-2000; 2000US-0211607P.
 PR (CLON-) CLONTECH LAB INC.
 PA Pradkov AF, Tersikh A;
 PI WPI; 2002-154595/20.
 XX N-PSDB; AAD28209.
 DR New fluorescent timer proteins comprising an emission spectrum that
 PT changes over time from a first wavelength to a second wavelength, useful
 PT for monitoring intracellular protein movement, translocation, trafficking
 PT or stability.
 XX Claim 5; Fig 3; 89pp; English.
 PS The invention relates to a fluorescent timer protein having an emission
 CC spectrum that changes over time after synthesis from a first wavelength
 CC to a second wavelength. The fluorescent timer proteins are useful in
 CC monitoring the activity of a promoter, determining the age of a protein,
 CC identifying an agent that modulates the activity of a promoter and in

enriching a population of cells comprising a fluorescent timer protein. The fluorescent timer proteins are also useful for assessing gene expression during development of a multicellular organism or during cellular differentiation, in response to a drug or other inducer of promoter activity, as a reporter to serve as a read-out of promoter activity, monitoring intracellular protein movement or translocation, protein trafficking, or protein stability, to investigate temporal aspects of the activity of a regulatory element, for determining cell fate during development and organ remodelling, in spatial and temporal visualisation of newly synthesised proteins and accumulated proteins, and in distinguishing between newly formed and pre-existing structures, e.g. membrane junctions and extracellular matrix components. The fluorescent timer proteins may further be used to investigations where photobleaching techniques are employed, as detectable labels, as selectable markers, as biosensors in prokaryotic and eukaryotic cells, in protease cleavage assays, and as second messenger detectors. The nucleic acids can be used to generate transgenic, non-human plants or animals or site-specific gene modifications in cell lines. The present sequence is *Drosophila* sp. E5NA non-aggregating fluorescent timer protein derived from fluorescent timer protein E5 by substituting amino acids at positions R2A, K5E and K9T. E5 protein is derived from humanised wild-type Anchozoa protein drFP583 by substituting Val to Ala at 105 and Ser to Thr at 197

	Pred. No.: Score: Percent Similarity: Best Local Similarity: Query Match: DB:	4,986-102 1214.00 100.00% 100.00% 98.14% 5	Length: Matches: Conservative: Mismatches: Indels: Gaps:	225 225 0 0 0 0
US-10-C8J1-864-14 (1-678) x AAE17542 (1-235)				

CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
 CC invention. Note: This sequence is not shown in the specification, but is
 CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
 CC ID NO:8 (AAE28833) in page 70-71 of the specification
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 4,08e-101 Length: 225
 Score: 1204.00 Matches: 223
 Percent Similarity: 99.11% Conservative: 0
 Best Local Similarity: 99.11% Mismatches: 2
 Query Match: 97.33% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x AAE28836 (1-225)

QY 1 ATGGCTCTCCCGAGAACGTCATCACCGAGTTCATGGCTTCAAGTGGCGATGAGGGC 60
 DB 1 MetAlaSerSerGluAsnValIleThrGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACGGCCACAGTTTCGAGATCGAGGGCGAGGGCGGCCGCCCTACGAGGGC 120
 DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACACACCGTGAAGTTGAAGGTGACCAAGGGGGCCCGCTCCCTCCCTGGGACATC 180
 DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCCCGCCAGTTCCAGTACGGCTCAAGTGTACGTGAAGCACCCCGCCGACATCCCC 240
 DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTACGAAGAGTCTCTCCCGAGGGCTTCAAGTGGAGGGCGCTGATGAATCTCGAG 300
 DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGGCGCTGGCGACCGGTGACCCAGGACTCTCCCTCGAGGACGGCTGCTCATCTAC 360
 DB 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGTGAAGTTCATCGCGGTAACTCTCCCTCGAGGGCCCGCTGATGAGAGAGACC 420
 DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 QY 421 ATGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGGACGGCTGTGAAGGCGAG 480
 DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATCCAAAGCCCTGAAGCTGAAGGACGGCGGACCTACTCTGAGTTCAGTCCATC 540
 DB 161 ThrHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGCCAAGAGCCGCGAGCTGCCCGCTACTACTAGTGCACACCAAGCTGCAC 600
 DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspAlaLysLeuAsp 200
 QY 601 ATCACTTCCCAACAGGAGTACTACCATCTGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
 DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACACCTGTTCCTG 675
 DB 221 HisHisLeuPheLeu 225

RESULT 4

ID AAE28921 standard; protein; 225 AA.

XX AC AAE28921;

XX

DT 27-DEC-2002 (first entry)
 XX Discosoma sp. drFP583 (NFP-6) mutant protein, E5 (V105A-S197T).
 DE Fluorescent protein; chromoprotein; protease cleavage assay; filter;
 XX fluorescence activated cell sorting application; fluorescent timer;
 KW biosensor; fluorescence resonance energy transfer application; FRET;
 KW colouring agent; recombinant DNA application; analyte detection assay;
 KW sunsreen; second messenger detector; drFP583 protein; NFP-6; mutant;
 KW muten.
 XX Discosoma sp.
 OS Synthetic.
 OS
 XX Key Location/Qualifiers
 FT Misc-difference 105 /note= "Wild-type Val substituted with Ala"
 FT Misc-difference 217 /note= "Wild-type Ser substituted with Thr"
 FT
 XX WO200268459-A2.
 XX 06-SEP-2002.
 XX 20-FEB-2002; 2002WO-US005749.
 XX 21-FEB-2001; 2001US-0270983P.
 XX 04-DEC-2001; 2001US-00006922.
 XX (CLON-) CLONTECH LAB INC.
 PA Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
 PI WPI; 2002-691654/74.
 XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 PT of an aggregating Chidarian chromo- or fluorescent protein or mutant for
 PT analyte detection assays or fluorescence activated cell sorting
 PT applications.
 PS Disclosure; Page; 80pp; English.
 CC The invention relates to nucleic acid molecules encoding non-aggregating
 CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
 CC invention. Note: This sequence is not shown in the specification, but is
 CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
 CC ID NO:8 (AAE28833) in page 70-71 of the specification
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 1.17e-100 Length: 225
 Score: 1199.00 Matches: 222
 Percent Similarity: 99.11% Conservative: 1
 Best Local Similarity: 98.67% Mismatches: 2
 Query Match: 96.93% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x AAE28921 (1-225)

QY 1 ATGGCTCTCCCGAGAACGTCATCACCGAGTTCATGGCTTCAAGTGGCGATGAGGGC 60
 DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGGCCGCCCTACGAGGGC 120

Db 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Qy 361 AAGGTGAAGTTCATCGCGGTGAACCTCCCTCCGACGCGCCCTGATGACAGAAAGACC 420
 Db 121 LysVallysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Qy 421 ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGCGACGCGGTGCTGAAGGCGGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Qy 481 ATCCACAGGCGCTGAAGCTGAAGGCGGCGCCACTTACCTGTGAGTCAAGTCCATC 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 Qy 541 TACATGCCCAAGAGCGCGTGCAGCTGCCCGGTACTACTACTGACACACCAAGCTGGAC 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
 Qy 601 ATCAGCTCCACAGGAGGACTACACCATCTGCGAGCTACGAGCGCACCGAGGCGCG 660
 Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Qy 661 CACCACCTGTTCTCTG 675
 Db 221 HisHisLeuPheLeu 225

RESULT 6
 AAE28925 standard; protein; 225 AA.
 XX AC AAE28925;
 XX DT 27-DEC-2002 (first entry)
 XX DE Discosoma sp. drFP583 (NFP-6) mutant protein, AG4 (V71M-V105A+S197T).
 XX KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
 XX KW biosensor; fluorescence resonance energy transfer application; FRET;
 XX KW colouring agent; recombinant DNA application; analyte detection assay;
 XX KW sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;
 XX KW mutin.
 OS Discosoma sp.
 OS Synthetic.
 PH Key Location/Qualifiers
 FT Misc-difference 71 /note= "Wild-type Val substituted with Met"
 FT FT Misc-difference 105 /note= "Wild-type Val substituted with Ala"
 FT FT Misc-difference 197 /note= "Wild-type Ser substituted with Thr"
 FT FT
 FN W0200268459-A2.
 XX
 XX 06-SEP-2002.
 XX 20-FEB-2002; 2002WO-US0005749.
 XX 21-FEB-2001; 2001US-0270983P.
 PR 04-DEC-2001; 2001US-00006922.
 XX
 XX (CLON-) CLONTECH LAB INC.
 XX Lukanov S, Lukanov K, Yanushevich Y, Savitsky A, Fradkov A;
 XX WPI; 2002-691654/74.
 XX
 XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 PT of an aggregating Chidarian chromo- or fluorescent protein or mutant for
 PT analyte detection assays or fluorescence activated cell sorting
 PT applications.

XX Disclosure; Page; 80pp; English.
 PS The invention relates to nucleic acid molecules encoding non-aggregating
 CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
 CC invention. Note: This sequence is not shown in the specification, but is
 CC derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
 CC ID NO:8 (AAE28925) in page 70-71 of the specification
 XX
 SQ Sequence 225 AA;
 Alignment Scores:
 Pred. No.: 2,19e-100 Length: 225
 Score: 1196.00 Matches: 221
 Percent Similarity: 99.11% Conservative: 2
 Best Local Similarity: 98.22% Mismatches: 0
 Query Match: 96.69% Indels: 0
 DB: 5 Gaps: 0
 US-10-081-864-14 (1-678) x AAE28925 (1-225)
 Qy 1 ATGCGCTCTCCGAGAACGTATCACCGAGTTCATCGCTTCAAGTGCATGAGGCG 60
 Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 Qy 61 ACCGTGAACGCGCAGGATTCGAGATCGAGGCGGCGAGGCGCGCCCTACGAGGCG 120
 Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGlyGly 40
 Qy 121 CACACACCGCTGAAGTGAAGGTGACCAAGGCGCGCCCTCGCTTCCGCTCGGACATC 180
 Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTyrAspIle 60
 Qy 181 CTGTCCCGCCGATTCAGTACGCTCAAGGTGTACGTGAAGCACCCCGCGCATCCTCC 240
 Db 61 LeuSerProGlnPheGlnTyrGlySerLysMetTyrValLysHisProAlaAspIlePro 80
 Qy 241 GACTACAAGAGCTGTCTTCCCGAGGCGTCAAGTGGGAGCGCGTGAATGAATTCGAG 300
 Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 Qy 301 GACGCGCGGTGGCGACCGGTGACCCAGGACTCTCTCCCTGAGGACGCGCTGCTCATCTAC 360
 Db 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Qy 361 AAGGTGAAGTTCATCGCGGTGAACCTCCCTCCGACGCGCCCTGATGACAGAAAGACC 420
 Db 121 LysVallysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Qy 421 ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGCGACGCGGTGCTGAAGGCGGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Qy 481 ATCCACAGGCGCTGAAGCTGAAGGCGGCGCCACTTACCTGTGAGTCAAGTCCATC 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 Qy 541 TACATGCCCAAGAGCGCGTGCAGCTGCCCGGTACTACTACTGACACACCAAGCTGGAC 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
 Qy 601 ATCAGCTCCACAGGAGGACTACACCATCTGCGAGCTACGAGCGCACCGAGGCGCG 660
 Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Qy 661 CACCACCTGTTCTCTG 675

CC sequence is constructed based on Discosoma sp. wild-type RFP protein
CC shown in page 30-31 (ABW00318)
XX
SQ Sequence 225 AA;

Alignment Scores:
Pred. No.: 2,7e-100 Length: 225
Score: 1195.00 Matches: 221
Percent Similarity: 98.67% Conservative: 1
Best Local Similarity: 98.22% Mismatches: 3
Query Match: 96.60% Indels: 0
DB: 7 Gaps: 0

US-10-081-864-14 (1-678) x ABW00937 (1-225)

QY 1 ATGGCTCTCCGAGAACGTCATCACCAGGTCATCGCGTTCAGAGTGGCGATGAGGCG 60
DB 1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTCAGAGTTCAGAGGCGAGGCGGCGCCCTACGAGGCG 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACACACCGTGAAGTTGAAGTGACCAAGGGCGGCGCCCTCCCTTCGCTGGGACATC 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCGCCAGTTCAGTAAGGCTCCCAAGGTGTACGTGAACACCGCCGACATCCCC 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValThrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAGACGTGTCCTCCCGAGGCGTTCAGAGTGGAGCGCGTGTATGAACCTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGTGGCGACCGTGAACCGAGGACTCTCCCTCGAGGCGGTGCTTCATCTAC 360
DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTCATCGCGGTGAACCTCCCTCCGACGCGCCCGTGTATGAGAGAGACC 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGGCGGCGGTGCTGAGGCGGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAAGGCGCTGAAGTGAAGGACGCGCGGCGCTACTCTGTGGAGTTCAGTCCATC 540
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCAAGACCGCTGAGTGTGCGGCTACTACTGTGGACACCAAGCTGGAC 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspThrLysLeuAsp 200
QY 601 ATCAGCTCCCAACAGGAGCTACACCATCGTGGAGCAGTACGAGGCGACCGAGGCGCG 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACTGTCCTG 675
DB 221 HisHisLeuPheLeu 225

RESULT 10
ID AAY99836
AA AAY99836 standard; protein; 225 AA.
XX
AC AAY99836;
XX
XX 12-SEP-2003 (revised)
DT 19-SEP-2000 (first entry)
XX
DE Discosoma sp. "red" novel fluorescent protein drFP583.

XX Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism;
KW fluorescent labeling.
XX
OS Discosoma sp; "red".
XX
FH Key Location/Qualifiers
FT Misc-difference 122
ET /note= "encoded by TC"
ET Misc-difference 127
ET /note= "encoded by GTTG"
XX
XX W0200034326-A1.
XX
XX 15-JUN-2000.
PD
XX 10-DEC-1999; 99WO-US029473.
DP
XX 11-DEC-1998; 98US-00210330.
PR
XX 14-OCT-1999; 99US-00418529.
PR
XX (CLON-) CLONTECH LAB INC.
PA
XX Lukyancy SA, Pradkov AF, Labas YA, Matz MV, Green G, Chen Y;
FI Ding L;
FI
DR WPI; 2000-423381/36.
XX
XX Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
XX useful for fluorescent labeling and as markers.
XX
XX Claim 20; Page 74-75; 86pp; English.
XX
XX The present sequence is a novel fluorescent protein (nFP) encoded by the
CC full-length cDNA drFP583. drFP583 was isolated from Discosoma sp. "red",
CC a non-bioluminescent species of the Class Anthozoa. Fluorescent proteins
CC can be used in fluorescent labeling, a useful tool for marking a protein,
CC cell or organism of interest. Unlike other markers used in protein
CC labeling, such as beta-galactosidase and luciferase, fluorescent proteins
CC do not require an exogenous cofactor or substrate. Methods involving
CC fluorescent proteins are also less laborious and less difficult to
CC control than the traditional methods of fluorescent labeling, where a
CC protein of interest is purified and then covalently conjugated to a
CC fluorophore derivative. Novel fluorescent proteins isolated from species
CC of the Class Anthozoa can be used as markers for gene expression and
CC protein localization studies, and in fluorescence resonance energy
CC transfer (FRET) reactions. They may have improved properties and better
CC suitability for larger excitations compared to prior art fluorescent
CC proteins such as green fluorescent protein. (Updated on 12-SEP-2003 to
CC standardise OS field)
XX
SQ Sequence 225 AA;

Alignment Scores:
Pred. No.: 6,26e-100 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 3 Gaps: 0

US-10-081-864-14 (1-678) x AAY99836 (1-225)

QY 1 ATGGCTCTCCGAGAACGTCATCACCAGTTCATCGCGTTCAGAGTGGCGATGAGGCG 60
DB 1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTCAGAGTTCAGAGGCGAGGCGGCGCCCTACGAGGCG 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACACACCGTGAAGTTCAGAGTTCAGAGGCGGCGGCGCCCTCCCTTCGCTGGGACATC 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCGCCAGTTCAGTAAGGCTCCCAAGGTGTACGTGAACACCGCCGACATCCCC 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValThrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAGACGTGTCCTCCCGAGGCGTTCAGAGTGGAGCGCGTGTATGAACCTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGTGGCGACCGTGAACCGAGGACTCTCCCTCGAGGCGGTGCTTCATCTAC 360
DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTCATCGCGGTGAACCTCCCTCCGACGCGCCCGTGTATGAGAGAGACC 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGGCGGCGGTGCTGAGGCGGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAAGGCGCTGAAGTGAAGGACGCGCGGCGCTACTCTGTGGAGTTCAGTCCATC 540
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCAAGACCGCTGAGTGTGCGGCTACTACTGTGGACACCAAGCTGGAC 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspThrLysLeuAsp 200
QY 601 ATCAGCTCCCAACAGGAGCTACACCATCGTGGAGCAGTACGAGGCGACCGAGGCGCG 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACTGTCCTG 675
DB 221 HisHisLeuPheLeu 225

Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheLeuTrpAspIle 60
 Qy 181 CTGCTCCCGCCAGTTCACAGTACGGCTCCCAAGCTGTAGCTGAAGCACCCCGCCGACATCCCC 240
 Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 Qy 241 GACTACAAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGAGCGCTGATGAATCTCGAG 300
 Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrPduLargValMetAsnPheGlu 100
 Qy 301 GACGGCGCTGGCGCGCGTACCGAGCTCTCCCTCGAGGAGCGCTGCTTCATCTAC 360
 Db 101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Qy 361 AAGGTGAAGTTCATCGCGCTGAATCTCCCTCCGACGGCGCCCTGATGAGAGAGAGACC 420
 Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Qy 421 ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGGACGGCTGTGAAGGCGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Qy 481 ATCCACAGGCGCTGAAGCTGAAGGAGCGGCGCCACTACTGCTGGAGTTCAGTCCATC 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 Qy 541 TACATGCCAGAGAGCGCTGAGCTGCGCGGCTACTACTACTGACACCAAGCTGGAC 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 Qy 601 ATCACCCTCCACACAGGAGCTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCG 660
 Db 201 IleThrSerHisGlnAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Qy 661 CACCACCTGTTCTCTG 675
 Db 221 HisHisLeuPheLeu 225

RESULT 11
 ID AAB01622 standard; protein; 225 AA.
 AC AAB01622;
 XX DT 12-DEC-2000 (first entry)
 XX DE Discosoma sp. red fluorescent protein drFP583.
 XX KW Anthozoa; fluorescent protein; fluorescence labeling; amFP486; cFP484;
 XX KM zFP506; zFP538; drFP583; dsFP483; asFP600; dgFP512; dmFP592.
 XX OS Discosoma sp.
 XX PN WO200034526-A1.
 XX PD 15-JUN-2000.
 XX PF 10-DEC-1999; 99WO-US029405.
 XX PR 11-DEC-1998; 98US-00210330.
 XX PA (CLON-) CLONTECH LAB INC.
 XX PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV;
 XX DR WPI; 2000-423451/36.
 XX PT Novel method for identifying a DNA sequence encoding fluorescent proteins
 XX PT from non-bioluminescent Anthozoa which are useful for fluorescent
 XX PT labeling and as markers.
 XX PS Claim 3; Page 68-69; 73pp; English.
 XX PS

CC The present sequence is Discosoma sp. red fluorescent protein drFP583. It
 CC was isolated using the Aequoria victoria green fluorescent protein (GFP)
 CC sequence, which was used to design PCR primers which might isolate other
 CC fluorescent proteins from a number of species of Anthozoa. These were
 CC Anemonia majano, Clavularia sp., Zoanthus sp., Discosoma sp. and Anemonia
 CC sulcata. The cDNA obtained was then screened in the search for sequences
 CC encoding fluorescent proteins. The other proteins found in this manner
 CC were cFP484, zFP506, zFP538, amFP486, dsFP483, asFP600, dgFP512 and
 CC dmFP592. These proteins can be used as fluorescent labels (for gene
 CC expression and protein localisation studies) and in fluorescence resonance
 CC energy transfer (FRET) studies in place of fluorophore derivatives and
 CC luciferases, as these involve laborious processes and the latter require
 CC cofactors. They can also be used in place of GFP, which is too stable to
 CC be useful when studying short-term or repetitive events
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 6,26e-100 Length: 225
 Score: 1191.00 Matches: 220
 Percent Similarity: 98.67% Conservatives: 2
 Best Local Similarity: 97.78% Mismatches: 3
 Query Match: 96.28% Indels: 0
 DB: 3 Gaps: 0

US-10-081-864-14 (1-678) x AAB01622 (1-225)

Qy 1 ATGGCTCTCCGAGAGCTCATCACCGAGTTCATGCGCTTCAAGTGCCTGAGGCG 60
 Db 1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 Qy 61 ACCGTGAAGCGCCAGGATTCGAGATCGAGGCGGAGCGGCGGCGCCCTCAGGAGGC 120
 Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 Qy 121 CACAACACCGTGAAGTTGAAGGTGACCAAGGCGGCGCCCTGCGCTTCGGGACATC 180
 Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 Qy 181 CTGTCCCGCCAGTTCAGTACGGCTCCAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240
 Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 Qy 241 GACTACAAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGAGCGCGTGTATGAACTTCGAG 300
 Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrPduLargValMetAsnPheGlu 100
 Qy 301 GACGGCGCTGGCGAGCGCTCCACGAGCGCTGTACCCCGGACCGCGCTGTCAATCTAC 360
 Db 101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Qy 361 AAGGTGAAGTTCATCGCGCTGAATCTCCCTCCGACGGCGCCCTGATGAGAGAGAGACC 420
 Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Qy 421 ATGGGCTGGAGGCTCCACGAGCGCTGTACCCCGGACCGCGCTGTGAAGGCGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Qy 481 ATCCACAGGCGCTGAAGCTGAAGGAGCGGCGCCACTACTACTGACACCAAGAGACC 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 Qy 541 TACATGCCAGAGAGCGCTGAGCTGCGCGGCTACTACTACTGAGACCAAGCTGGAC 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 Qy 601 ATCACCCTCCACACAGGAGCTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCG 660
 Db 201 IleThrSerHisGlnAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Qy 661 CACCACCTGTTCTCTG 675

Db 221 HisHisLeuPheLeu 225

RESULT 12
AAG65509
ID AAG65509 standard; protein; 225 AA.

XX
AC AAG65509;
XX
DT 30-NOV-2001 (first entry)
XX
DE Anthozoan red fluorescent protein sequence.
XX
KW Anthozoan; Anthozoan; fluorescence; marker; FRET; red.
XX
OS Anthozoa.
XX
PN WO200162919-A1.
XX
XX 30-AUG-2001.
XX
PF 13-FEB-2001; 2001WO-US004625.
XX
PR 23-FEB-2000; 2000US-0184732P.
XX
PA (AURO-) AURORA BIOSCIENCES CORP.
XX
PI Nelson D, Zamaira E, Tsien R;
XX
DR WPI; 2001-557704/62.
XX
XX
PT Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise
functional red fluorescent proteins, and the encoding nucleic acids, with
key mutations for improving the proteins function.
XX
PS Disclosure; Page 85; 90pp; English.
XX
CC The invention provides a nucleic acid encoding functional red fluorescent
CC protein (II) that differs from the sequence of an Anthozoan red
CC fluorescent protein by at least one amino acid substitution, and with
CC different fluorescent properties. The red fluorescent protein of the
CC invention can be expressed by standard recombinant methodology. (II) are
CC used a fluorescent markers and FRET partners. It is used for identifying
CC protein-protein interactions. (II) is also suitable for multiplexed
CC fluorescent analysis and FRET-based applications using existing Aequeora
CC fluorescent proteins. (II) has improved brightness, reduced spectral
CC cross talk, and is rapidly and efficiently expressed in mammalian cells.
CC The key mutations in the encoding nucleic acids provide improved folding,
CC brightness, and create (II) with sharper, more defined excitation and
CC emission peaks when expressed in mammalian cells. The present sequence
CC represents an anthozoan fluorescent protein
XX
SQ Sequence 225 AA;

Alignment Scores:
Pred. No.: 6, 26e-100 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 4 Gaps: 0

US-10-081-864-14 (1-678) x AAG65509 (1-225)

QY 1 ATGGCCTCTCCGAGAACGTCATCACCGAGTTTCATCGCTTCAAGTGGCCATGGAGGCG 60
Db 1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTTCGAGATCGAGGGCGAGGGCGCGCCCTACGAGGGCG 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTGAAGGTGAACCAAGGGCGGCGCCCTGCGCTTCGGGCATC 180

Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCGCCAGTTCCAGTACGCTCCAGGTGTAGTGAAGCAGCCCGCGCATCCCC 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAGAAGCTGCTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATCTCGAG 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGCGCGGTGGCGACCGTGAACCGAGGACTCTCCCTGCGAGGCGGCTGCTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrValGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTTCATCGCGGTGAACCTTCCCTCCGACCGCGCCCTGATGCAGAGAAGACC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGACCGCGCTGCTGAAGGGCGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAAGGCCCTGAAGCTGAAGGACGGCGGCGCACTACCTGTTGGAGTTCAAGTCCATC 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCCAAGAAGCCGCTGCTGAGCTCCCGGCTTACTACTGCTGACACCAAGCTGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCAGCTCCCAACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGCG 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCCTG 675
Db 221 HisHisLeuPheLeu 225

RESULT 13
ABB08834
ID ABB08834 standard; protein; 225 AA.

XX
AC ABB08834;
XX
DT 29-MAY-2002 (first entry)
XX
DE Yeast optimised RFP SEQ ID NO 17.
XX
KW Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW Escherichia coli; green fluorescent protein; biotechnology.
XX
OS Anthozoa.
XX
PN DE20001395-U1.
XX
PD 15-MAR-2001.
XX
PF 27-JAN-2000; 2000DE-02001395.
XX
PR 27-JAN-2000; 2000DE-02001395.
XX
PA (GPCB-) GPC BIOTECH AG.
XX
XX WPI; 2002-228394/29.
DR N-PSDB; ABA95905, ABA95921, ABA95922.
XX
PT New DNA encoding red fluorescent protein, useful as marker in
PT biotechnology, has sequence optimized for expression in eukaryotes,
PT especially yeast or plants.
XX
PS Disclosure; Page 13-14; 19pp; German.
XX

CC The invention relates to DNA (I) containing either sequence ABA95905 or
 CC sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
 CC (yRFP). (I) are used to express red fluorescent protein (RFP) in
 CC eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
 CC plants, especially dicotyledonous plants including Nicotiana tabacum or
 CC Arabidopsis thaliana and also in prokaryotes, especially bacteria,
 CC especially Escherichia coli. RFP is useful in the same way as green
 CC fluorescent protein but is more generally applicable in modern
 CC biotechnology. (I) are optimised for expression in yeast and so generate
 CC RFP at higher levels with stronger fluorescence and thus lowers the
 CC detection limit and gives a better signal-to-noise ratio. The present
 CC sequence is that of the yeast optimised RFP
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 6.26e-100 Length: 225
 Score: 1191.00 Matches: 220
 Percent Similarity: 98.57% Conservative: 2
 Best Local Similarity: 97.78% Mismatches: 3
 Query Match: 96.28% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x ABB08834 (1-225)

QY 1 ATGGCTCTCCGAGAACGTCATCACCGAGTTTCATGGCTTCAAGGTGCGCATGGAGGC 60
 DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACCTGAAACGCCACGAGTTCGAGATCGAGGGGAGGGCGGCGCCCTACGAGGC 120
 DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACAACACCGTGAAGTTGAAGTGACCAAGGGGGCCCTGCGCTTCGCTCGGACATC 180
 DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCCTCCCGAGTTCAGGTACGGTCCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCC 240
 DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTACAAAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGAGGCGGTGATCACTTCGAG 300
 DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGGCGGCTGGCGACCGTACCGAGGACTCTCTCCGAGGACGCGCTGCTTCATCTAC 360
 DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGGTGAAGTTCATCGCGGTGAATCTCCCTCCGACGCGCCCTGTATGCGAAGAAGACC 420
 DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 QY 421 ATGGCTGGGAGGCTCCACGAGCGCTGTACCCCGAGCGGCTGTGAAGGCGGAG 480
 DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATCCAAAGCCCTGAAGTGAAGAGCGGCGGCACTACTGCTGAGTTCAGTCCATC 540
 DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGCCCAAGAGCCCGTGCAGCTGCCCGGTACTACTACGTGACACCAAGCTGCAC 600
 DB 181 TyrMetAlaLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
 QY 601 ATCACTCCCAACACGAGGACTACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
 DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACCACCTGTCTCTG 675
 DB 221 HisHisLeuPheLeu 225

RESULT 14

AAE28926
 ID AAE28926 standard; protein; 225 AA.

XX AAE28926;

DT 27-DEC-2002 (first entry)

XX Discosoma sp. drFP583 (NFP-6) mutant protein, (V71M+V105A+Y120H+S197T).
 DE Fluorescent protein; chromoprotein; protease cleavage assay; filter;
 KW fluorescence activated cell sorting application; fluorescent timer;
 KW biosensor; fluorescence resonance energy transfer application; FRET;
 KW colouring agent; recombinant DNA application; analyte detection assay;
 KW sunscreen; second messenger detector; drFP583 protein; NFP-6; mutant;
 KW mutin.

XX Discosoma sp.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 71 /note= "Wild-type Val substituted with Met"

FT Misc-difference 105 /note= "Wild-type Val substituted with Ala"

FT Misc-difference 120 /note= "Wild-type Tyr substituted with His"

FT Misc-difference 197 /note= "Wild-type Ser substituted with Thr"

PN WO200268459-A2.

XX 06-SEP-2002.

XX 20-FEB-2002; 2002WO-US005749.

XX 21-FEB-2001; 2001US-0270983P.

XX 04-DEC-2001; 2001US-00006922.

PA (CLON-) CLONTECH LAB INC.

XX Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Pradkov A;
 PI WPI; 2002-691654/74.

XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
 an analyte detection assays or fluorescence activated cell sorting
 applications.

XX Disclosure; Page: 80pp; English.

XX The invention relates to nucleic acid molecules encoding non-aggregating
 chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 useful in analyte detection assays, as colouring agents, as markers in
 recombinant DNA applications, as sunscreens or filters, in fluorescence
 resonance energy transfer (FRET) applications, as biosensors in
 prokaryotic and eukaryotic cells, in screening assays, as second
 messenger detectors, in fluorescence activated cell sorting applications,
 in protease cleavage assays or as fluorescent timers. The present
 sequence is Discosoma sp. drFP583 (NFP-6) mutant protein of the
 invention. Note: This sequence is not shown in the specification, but is
 derived from Discosoma sp. drFP583 (NFP-6) wild-type protein shown as SEQ
 ID NO:8 (AAE288833) in page 70-71 of the specification

XX Sequence 225 AA;

Alignment Scores:

Pred. No.: 6.26e-100 Length: 225
 Score: 1191.00 Matches: 220
 Percent Similarity: 99.11% Conservative: 3
 Best Local Similarity: 97.78% Mismatches: 2
 Query Match: 96.28% Indels: 0

XX 20-FEB-2002; 2002WO-US005749.
XX PF
XX XX
PR 21-FEB-2001; 2001US-0270983P.
PR 04-DEC-2001; 2001US-00006922.
XX (CLON-) CLONTECH LAB INC.
XX LUKYANOV S, LUKYANOV K, YANUSHEVICH Y, SAVITSKY A, FRADKOV A;
XX WPI; 2002-691654/74.
XX N-PSDB; AD46278.
XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
XX of an aggregating Chidarian chromo- or fluorescent protein or mutant for
XX analyte detection assays or fluorescence activated cell sorting
XX applications.
XX
XX Disclosure; Page 70-71; 80pp; English.
XX
XX The invention relates to nucleic acid molecules encoding non-aggregating
XX chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
XX useful in analyte detection assays, as colouring agents, as markers in
XX recombinant DNA applications, as sunscreens or filters, in fluorescence
XX resonance energy transfer (FRET) applications, as biosensors in
XX prokaryotic and eukaryotic cells, in screening assays, as second
XX messenger detectors, in fluorescence activated cell sorting applications,
XX in protease cleavage assays or as fluorescent timers. The present
XX sequence is Discosoma sp. drFP583 (NFP-6) wild-type protein of the
XX invention
XX
XX Sequence 225 AA;
SQ
Alignment Scores:
Pred. No.: Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: Gaps: 5
US-10-081-864-14 (1-678) x AAE28833 (1-225)
QY 1 ATGGCTCTCTCCGAGAACGTCATCCAGGTTTCATGCGCTTCAAGGTGCGATGAGGCG 60
DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGGCGCCCTTACGAGGCG 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAAACCGCTGAAGTTCGAGTCCAGGCGGCGCCCTTCCCTGCGGACATC 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAsp 60
QY 181 CTGTCTCCCGGAGTTCAGTTCAGGTCCTCAAGGTGACGAGCGGCGGCGGCGGCGG 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAsp 80
QY 241 GACTACAGAGAGTCTCTCTCCCGAGGCTTCAAGTGGGAGCGGCTGATGAATTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCGTGGCGAGCCGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTCATCGGCGTGAACCTCCCTCCGAGGCGGCGGCGGCGGCGGAGGAG 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysThr 140
QY 421 ATGGGCTGGGAGGCGCTCCACCGAGGCGCTGTACCCCGCGGAGCGGCGTGTGAAGGCGGAG 480

DB: 5 Gaps: 0
US-10-081-864-14 (1-678) x AAE28833 (1-225)
QY 1 ATGGCTCTCTCCGAGAACGTCATCCAGGTTTCATGCGCTTCAAGGTGCGATGAGGCG 60
DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGGCGCCCTTACGAGGCG 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAAACCGCTGAAGTTCGAGTCCAGGCGGCGCCCTTCCCTGCGGACATC 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAsp 60
QY 181 CTGTCTCCCGGAGTTCAGTTCAGGTCCTCAAGGTGACGAGCGGCGGCGGCGGCGG 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysMetTyrValLysHisProAlaAsp 80
QY 241 GACTACAGAGAGTCTCTCTCCCGAGGCTTCAAGTGGGAGCGGCTGATGAATTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCGTGGCGAGCCGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleHis 120
QY 361 AAGGTGAAGTTCATCGGCGTGAACCTCCCTCCGAGGCGGCGGCGGCGGCGGAGGAG 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysThr 140
QY 421 ATGGGCTGGGAGGCGCTCCACCGAGGCGCTGTACCCCGCGGAGCGGCTGTGAAGGCGGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGCCCTGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGCCCAAGAGCCGCTGAGCTGCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
QY 601 ATACCTCCACACAGGAGGAGTACCATCGTGGAGGAGTACGAGGAGGAGGAGGAGGAG 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCTCTG 675
DB 221 HisHisLeuPheLeu 225
RESULT 15
AAE28833
ID AAE28833 standard; protein; 225 AA.
AC AAE28833;
XX
XX 27-DEC-2002 (first entry)
DT
DE Discosoma sp. drFP583 (NFP-6) wild-type protein.
XX
XX Fluorescent protein; chromoprotein; protease cleavage assay; filter;
XX fluorescence activated cell sorting application; fluorescent timer;
XX biosensor; fluorescence resonance energy transfer application; FRET;
XX colouring agent; recombinant DNA application; analyte detection assay;
XX sunscreen; second messenger detector; drFP583 protein; NFP-6.
OS Discosoma sp.
XX
XX WO200268459-A2.
XX
XX 06-SEP-2002.
PD

Db	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
QY	481	ATCCACAAGGCCCTGAAGCTGAAGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATC	540
Db	161	IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
QY	541	TACATGGCCAAAGACCCGTCAGCTGCCCGCTACTACTACGTGGACACCAAGCTGGAC	600
Db	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp	200
QY	601	ATCACCTCCCAACAGGAGCTACACCATCGTGGAGCAGTACGCGACCGAGGCGCGC	660
Db	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
QY	661	CACCACCTGTTCTCTG	675
Db	221	HisIleuPheLeu	225

Search completed: July 29, 2004, 14:32:26
Job time : 51.2188 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:29:01 ; Search time 13.9516 Seconds

(without alignments)

5017.682 Million cell updates/sec

Title:

US-10-081-864-14

Perfect score:

1237

Sequence:

1 atggctctctcgagacgt.....gccaccacctgtctctgttaa 678

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10081864/runat_29072004_150507_25519/app_query.fasta_1.1742
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -ENDS=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10081864 -CGN_1_1_45 @runat_29072004_150507_25519 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA.*
1: /cgn2_6/prodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	475	38.4	238	3	US-09-277-716-16
2	475	38.4	238	4	US-09-609-161B-16
3	475	38.4	238	4	US-09-626-581D-65
4	475	38.4	238	4	US-09-415-765B-65
5	475	38.4	238	4	US-09-626-580C-65
6	463	37.4	238	4	US-09-839-650-3
7	462	37.3	238	3	US-09-277-716-32
8	462	37.3	238	4	US-09-609-161B-32
9	229.5	18.6	238	4	US-09-023-946B-23
10	228.5	18.5	238	4	US-09-023-946B-28
11	226.5	18.3	238	2	US-08-818-604-32
12	226.5	18.3	238	3	US-08-819-612-22

13	226.5	18.3	238	4	US-09-316-919-3	Sequence 3, Appl
14	226.5	18.3	238	4	US-09-346-946-32	Sequence 32, Appl
15	226.5	18.3	238	4	US-09-023-946B-22	Sequence 22, Appl
16	224.5	18.1	238	1	US-08-337-915A-2	Sequence 2, Appl
17	224.5	18.1	238	1	US-08-753-143-2	Sequence 2, Appl
18	224.5	18.1	238	2	US-08-679-865-2	Sequence 2, Appl
19	224.5	18.1	238	2	US-08-680-876-2	Sequence 2, Appl
20	224.5	18.1	238	2	US-08-792-553-2	Sequence 2, Appl
21	224.5	18.1	238	3	US-08-911-825-2	Sequence 2, Appl
22	224.5	18.1	238	3	US-08-753-144-2	Sequence 2, Appl
23	224.5	18.1	238	3	US-08-974-737-2	Sequence 2, Appl
24	224.5	18.1	238	3	US-08-706-408-2	Sequence 2, Appl
25	224.5	18.1	238	3	US-09-094-359-2	Sequence 2, Appl
26	224.5	18.1	238	3	US-09-172-063-2	Sequence 2, Appl
27	224.5	18.1	238	3	US-09-121-539-1	Sequence 1, Appl
28	224.5	18.1	238	3	US-09-263-975-2	Sequence 2, Appl
29	224.5	18.1	238	4	US-08-737-452-2	Sequence 2, Appl
30	224.5	18.1	238	4	US-09-465-142-2	Sequence 2, Appl
31	224.5	18.1	238	4	US-09-418-785-1	Sequence 1, Appl
32	224.5	18.1	238	4	US-09-214-509-2	Sequence 2, Appl
33	224.5	18.1	238	4	US-09-479-645A-10	Sequence 10, Appl
34	224.5	18.1	238	4	US-09-479-645A-159	Sequence 159, App
35	224.5	18.1	238	4	US-09-129-192C-2	Sequence 2, Appl
36	224.5	18.1	238	4	US-09-129-192C-74	Sequence 74, Appl
37	224.5	18.1	238	4	US-09-575-847-2	Sequence 2, Appl
38	224.5	18.1	238	4	US-09-602-641-2	Sequence 2, Appl
39	224.5	18.1	238	4	US-09-704-463-2	Sequence 2, Appl
40	224.5	18.1	238	4	US-09-472-065A-4	Sequence 4, Appl
41	224.5	18.1	238	4	US-09-603-448-7	Sequence 7, Appl
42	224.5	18.1	238	4	US-09-603-448-24	Sequence 24, Appl
43	224.5	18.1	238	4	US-09-023-946B-24	Sequence 24, Appl
44	224.5	18.1	238	4	US-09-920-922-4	Sequence 4, Appl
45	224.5	18.1	238	5	PCT-US95-14692-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-277-716-16
; Sequence 16, Application US/09277716A

; Patent No. 6232107

; GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce

; APPLICANT: Szent-Gyorgyi, Christopher

; APPLICANT: PROLUME, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE

; CURRENT APPLICATION NUMBER: US/09/277,716A

; CURRENT FILING DATE: 1999-03-26

; EARLIER APPLICATION NUMBER: 60/102,939

; EARLIER FILING DATE: 1998-10-01

; EARLIER APPLICATION NUMBER: 60/089,367

; EARLIER FILING DATE: 1998-06-15

; EARLIER APPLICATION NUMBER: 60/079,624

; EARLIER FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 16

; LENGTH: 238

; TYPE: PRT

; ORGANISM: Renilla mulleri

; FEATURE:

; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)

US-09-277-716-16

Alignment Scores:

Pred. No.: 2,24e-34 Length: 238
Score: 475.00 Matches: 93
Percent Similarity: 64.59% Conservative: 42
Best Local Similarity: 44.50% Mismatches: 72
Query Match: 38.40% Indels: 2
DB: 3 Gaps: 2

US-10-081-864-14 (1-678) x US-09-277-716-16 (1-238)

```

QY 22 ATCAGCGAGTTCATCGCTCAAGTGGCGATGGAGGCGACCGTGAACGCCACGAGTTC 81
DB 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyLeuValAsnHisValPhe 30
QY 82 GAGATCGAGGCGGAGGCGGCGCCCTACGAGGCGCACCAACACCGTGAAGTTGAAG 141
DB 31 ThrMetGluGlyCysGlyLysGlyAsnLeuPheGlyAsnGlnLeuValGlnileArg 50
QY 142 GTGACCAAGGCGGCGCCCTCGCTCGCTGGGACATCTGCTCCCGGAGTTCAGTAC 201
DB 51 ValThrLysGlyAlaProLeuProPheAlaPheAspLeuValSerProAlaPheGlnTyr 70
QY 202 GGCTCCAAAGTGTACGTGAAGACCCCGCGACATCTCCCTCGGACATCTCCCGGAGTTC 261
DB 71 GlyAsnArgThrPheThrLysTyrProAsnAspLeuValSerProAlaPheGlnTyr 90
QY 262 CCCAGGAGTTCATCGCTCAAGTGGCGATGATGATGATGATGATGATGATGATGATG 321
DB 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyLeuValGluile 110
QY 322 ACCGAGGAGTTCATCGCTCAAGTGGCGATGATGATGATGATGATGATGATGATG 381
DB 111 ArgSerAspLeuAsnLeuLeuGluAspLysPheValTyrArgValGluTyrLysGlySer 130
QY 382 AACTTCCCTCCGAGCGCGCGTGTATGACAGAGACCATGGCTGGGAGCGCTCCACC 441
DB 131 AsnPheProAspGlyProValMetGlnLysThrileLeuGlyLeuGluProSerPhe 150
QY 442 GAGCGCTGTACCCCGCGAGCGGTCTGACAGGCGGAGATCCACAGGCGCTGAAGCTG 501
DB 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGluValLeuValTyrLysLeu 170
QY 502 AAGGACGCGCGCCACTACCTCGTGGAGTTCAGTCCATCTACATGCGCAGAGCCCGTG 561
DB 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal 190
QY 562 ---CAGTGGCGGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 618
DB 191 LysGluPheProSerTyrHisPheIleGlnHisArgLeuGluLysThr---TyrValGlu 209
QY 619 GACTACACCATCGTGGAGCAGTACGAG 645
DB 210 AspGlyGlyPheValGluGlnHisGlu 218

RESULT 2
US-09-609-161B-16
; Sequence 16, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR FILING DATE: 1999-03-26
; PRIOR FILING DATE: 1999-03-26
; PRIOR FILING DATE: 1998-10-01
; PRIOR FILING DATE: 1998-10-01
; PRIOR FILING DATE: 1998-06-15
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri

```

```

; FEATURE:
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-16

Alignment Scores:
Pred. No.: 2,24e-34 Length: 238
Score: 475.00 Matches: 93
Percent Similarity: 64.59% Conservative: 42
Best Local Similarity: 44.50% Mismatches: 72
Query Match: 38.40% Indels: 2
DB: 4 Gaps: 2

US-10-081-864-14 (1-678) x US-09-609-161B-16 (1-238)
QY 22 ATCAGCGAGTTCATCGCTCAAGTGGCGATGGAGGCGACCGTGAACGCCACGAGTTC 81
DB 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyLeuValAsnHisValPhe 30
QY 82 GAGATCGAGGCGGAGGCGGCGCCCTACGAGGCGCACCAACACCGTGAAGTTGAAG 141
DB 31 ThrMetGluGlyCysGlyLysGlyAsnLeuPheGlyAsnLeuValGlnileArg 50
QY 142 GTGACCAAGGCGGCGCCCTCGCTCGCTGGGACATCTCCCTCGGACATCTCCCGGAGTTC 201
DB 51 ValThrLysGlyAlaProLeuProPheAlaPheAspLeuValSerProAlaPheGlnTyr 70
QY 202 GGCTCCAAAGTGTACGTGAAGACCCCGCGACATCTCCCGGAGTTCAGTAC 261
DB 71 GlyAsnArgThrPheThrLysTyrProAsnAspLeuValSerProAlaPheGlnTyr 90
QY 262 CCCAGGAGTTCATCGCTCAAGTGGCGATGATGATGATGATGATGATGATGATG 321
DB 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyLeuValGluile 110
QY 322 ACCGAGGAGTTCATCGCTCAAGTGGCGATGATGATGATGATGATGATGATGATG 381
DB 111 ArgSerAspLeuAsnLeuLeuGluAspLysPheValTyrArgValGluTyrLysGlySer 130
QY 382 AACTTCCCTCCGAGCGCGCGTGTATGACAGAGACCATGGCTGGGAGCGCTCCACC 441
DB 131 AsnPheProAspGlyProValMetGlnLysThrileLeuGlyLeuGluProSerPhe 150
QY 442 GAGCGCTGTACCCCGCGAGCGGTCTGACAGGCGGAGATCCACAGGCGCTGAAGCTG 501
DB 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGluValLeuValTyrLysLeu 170
QY 502 AAGGACGCGCGCCACTACCTCGTGGAGTTCAGTCCATCTACATGCGCAGAGCCCGTG 561
DB 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal 190
QY 562 ---CAGTGGCGGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 618
DB 191 LysGluPheProSerTyrHisPheIleGlnHisArgLeuGluLysThr---TyrValGlu 209
QY 619 GACTACACCATCGTGGAGCAGTACGAG 645
DB 210 AspGlyGlyPheValGluGlnHisGlu 218

RESULT 3
US-09-626-581D-65
; Sequence 65, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR FILING DATE: 1998-10-08
; PRIOR FILING DATE: 1998-10-08
; PRIOR FILING DATE: 1999-10-08

```


; FILE REFERENCE: A-66900-2/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/626,580C
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
; US-09-626-580C-65

Alignment Scores:
Pred. No.: 2,24e-34 Length: 238
Score: 475.00 Matches: 93
Percent Similarity: 64.59% Conservative: 42
Best Local Similarity: 44.50% Mismatches: 72
Query Match: 38.40% Indels: 2
DB: 4 Gaps: 2

US-10-081-864-14 (1-678) x US-09-626-580C-65 (1-238)

```
QY 22 ATACCGAGTTCATCGCGCTTCAGGTGCGCATCGAGGCGCACCGTGAACGGCCACGAGTTC 81
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyIleValAsnAsnHisValPhe 30
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 82 GAGATCAGGCGGAGGCGGAGGCGCGCCCTACGAGGCGCACCAACACCGCTGAAGTTGAAG 141
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 31 ThrMetGluGlyCysGlyLysGlyAsnIleLeuPheGlyAsnGlnLeuValGlnIleArg 50
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 142 GTGACCAAGGCGGCGCCCTCGCTGCGTGGGACATCTGCTCCCGCCAGTTCAGTAC 201
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 51 ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 202 GGTCTCAAGGTGTACGTGAAGCACCGCGCGACATCCCGACATACAAAGAGTGTCTCTC 261
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 262 CCCGAGGCGTCAAGTGGGAGCGCGTGAATCTGAGAGCGCGCGTGGCGACCGTG 321
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 91 ProAlaGlyPheMetTyrGluAArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 322 ACCGAGACTCTCCCTCGCAGGACGCGTCTCATCAAGGTGAAGTTCATCGCGGTG 381
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 111 ArgSerAspIleAsnLeuIleGluAspLysPheValTyrArgValGluTyrLysGlySer 130
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 382 AACTTCCCTCCGACGCGCCCGTGTATGCAGAGAAGACCATGGCTGGGAGCGCTCCACC 441
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 442 GAGGCGCTGTACCCCGCAGCGCGTGTGAAGGCGGAGATCCACAAGGCCCTGAAGTG 501
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluValIleLeuValTyrLysLeu 170
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 502 AAGGACGCGCGCCACTACCTGTTGGAGTTCAAGTTCATCTACATGCGCCAAAGACCGGTG 561
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal 190
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 562 ---CAGTGGCGCGCTACTACTAGTGGAGACCAACAGCTGACATCACTCCCAACAGAG 618
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 191 LysGluPheProSerTyrHisPheIleGlnHisArgLeuGluLysThr---TyrValGlu 209
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 619 GACTACACCATCGTGGAGCAGTACGAG 645
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 210 AspGlyGlyPheValGluGlnHisGlu 218
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

RESULT 6

US-09-626-580-3
; Sequence 3, Application US/09839650
; Patent No. 6645761

; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; TITLE OF INVENTION: Fluorescent Protein
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
; US-09-839-650-3

Alignment Scores:
Pred. No.: 2,63e-33 Length: 238
Score: 463.00 Matches: 92
Percent Similarity: 63.64% Conservative: 41
Best Local Similarity: 44.02% Mismatches: 74
Query Match: 37.43% Indels: 2
DB: 4 Gaps: 2

US-10-081-864-14 (1-678) x US-09-839-650-3 (1-238)

```
QY 22 ATACCGAGTTCATCGCGCTTCAGGTGCGCATCGAGGCGCACCGTGAACGGCCACGAGTTC 81
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyIleValAsnAsnHisValPhe 30
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 82 GAGATCAGGCGGAGGCGGAGGCGCGCCCTACGAGGCGCACCAACACCGCTGAAGTTGAAG 141
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 31 ThrMetGluGlyCysArgLysArgAsnIleLeuPheGlyAsnGlnLeuValHisIleArg 50
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 142 GTGACCAAGGCGGCGCCCTCGCTGCGTGGGACATCTGCTCCCGCCAGTTCAGTAC 201
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 51 ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 202 GGTCTCAAGGTGTACGTGAAGCACCGCGCGACATCCCGACATACAAAGAGTGTCTCTC 261
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 262 CCCGAGGCGTCAAGTGGGAGCGCGTGAATCTGAGAGCGCGCGTGGCGACCGTG 321
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 91 ProAlaGlyPheMetTyrGluAArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 322 ACCGAGACTCTCCCTCGCAGGACGCGTCTCATCAAGGTGAAGTTCATCGCGGTG 381
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 111 ArgSerAspIleAsnLeuIleGluAspLysPheValTyrArgValGluTyrLysGlySer 130
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 382 AACTTCCCTCCGACGCGCCCGTGTATGCAGAGAAGACCATGGCTGGGAGCGCTCCACC 441
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 442 GAGGCGCTGTACCCCGCAGCGCGTGTGAAGGCGGAGATCCACAAGGCCCTGAAGTG 501
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluValIleLeuValTyrLysLeu 170
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 502 AAGGACGCGCGCCACTACCTGTTGGAGTTCAAGTTCATCTACATGCGCCAAAGACCGGTG 561
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLysGlyValVal 190
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 562 ---CAGTGGCGCGCTACTACTAGTGGAGACCAACAGCTGACATCACTCCCAACAGAG 618
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 191 LysGluPheProSerTyrHisPheIleGlnHisArgLeuGluLysThr---TyrValGlu 209
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 619 GACTACACCATCGTGGAGCAGTACGAG 645
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 210 AspGlyGlyPheValGluGlnHisGlu 218
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

RESULT 7

US-09-277-716-32
; Sequence 32, Application US/09277716A

Db 131 GlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMetGluProSerPhe 150
 Qy 442 GAGCGCTGTACCCCGGACGGCTGCTGAAGGGGAGATCCACAAAGCCCTGAAGCTG 501
 Db 151 GluValValTyrMetAsnSerGlyValLeuValGluValAspLeuValTyrLysLeu 170
 Qy 502 AAGGACGGCGCCACTACCTGCTGGAGTTCAAGTCCATCTACATGCCAGAGCCCGTG 561
 Db 171 GluSerGlyAsnTyrTyrSerCysHisMetLysThrPheTyrArgSerLysGlyVal 190
 Qy 562 ---CAGTGGCCGCTACTACTAGTGGACACCAAGCTGCATCATCTCCCAACAGAG 618
 Db 191 LysGluPheProGluTyrHisPheIleHisArgLeuGluLysThr---TyrValGlu 209
 Qy 619 GACTACACCATCGTGGAGCAGTACGAG 645
 Db 210 GluGlySerPheValGluGlnHisGlu 218

RESULT 9

US-09-023-946B-23
 ; Sequence 23 Application US/09023946B
 ; Patent No. 6670449

GENERAL INFORMATION:

APPLICANT: GERO MIESENBOCK, ET AL.
 TITLE OF INVENTION: HYBRID MOLECULES AND
 THEIR USE FOR OPTICALLY DETECTING CHANGES IN

CELLULAR MICROENVIRONMENTS

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/023,946B
 FILING DATE: 13-FEB-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/036,805
 FILING DATE: 14 FEBRUARY 1997
 APPLICATION NUMBER: 60/038,179
 FILING DATE: 13 FEBRUARY 1997

ATTORNEY/AGENT INFORMATION:

NAME: KENNETH H. SONNENFELD
 REGISTRATION NUMBER: 33,285
 REFERENCE/DOCKET NUMBER: 2955-4004US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800
 TELEFAX: (212) 751-5849
 TELE: 421792

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:
 LENGTH: 238
 TYPE: AMINO ACID
 STRANDEDNESS: UNKNOWN
 TOPOLOGY: UNKNOWN

SEQUENCE DESCRIPTION: SEQ ID NO: 23:

US-09-023-946B-23

Alignment Scores:

Pred. No.:	1.76e-12	Length:	238
Score:	229.50	Matches:	60
Percent Similarity:	47.21%	Conservative:	50
Best Local Similarity:	25.75%	Mismatches:	110
Query Match:	18.55%	Indels:	13
DB:	4	Gaps:	6

US-10-081-864-14 (1-678) x US-09-023-946B-23 (1-238)
 Qy 1 ATGGCTCTCTCCGAGAACGTTCATCACCAGAGTTTCATGCGCTTCAAGTGGCCATGAGGGC 60
 Db 1 MetSerLysGlyGluGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
 Qy 61 ACCGTGAACGCCCGAGTTCGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGC 120
 Db 21 AspValAsnGlyHisLysPheSerValSerGlyGluGlyAspAlaThrTyrGly 40
 Qy 121 CACAACACCGTGAAGTTGAAGTGACCAAGGGCGGCCCTTCGCTTCCGCTGGGACATC 180
 Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProTyrProThr 59
 Qy 181 CTGTCCCGCCAGTTCAGTACGGCTCCAGAGTGTACGTGAAGCACCCCGCGCATCCCC 240
 Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
 Qy 241 -----GACTACAAGAAGCTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAAC 294
 Db 80 ArgHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluA:GThrIlePhe 99
 Qy 295 TTCGAGGACGGCGGCGTGGCGACCGTGCACCCAGGACTCTCCCTGCAGGACGGCTGCTC 354
 Db 100 PheLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
 Qy 355 ATCTACAAGGTGAAGTTTCATCGCGTGAAGTTCCTCCCGAGGGCTTCCCGAGGGCGCGTGAATGAAG 414
 Db 120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 139
 Qy 415 AAGACCATGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGAGGGGCTGCTGAAG 474
 Db 140 Lys---LeuGluTyrAsnTyrAsnAspHisGlnValTyrIleMetAlaAspLysGlnLys 158
 Qy 475 GCGGAGATCCACAAGGCCCTGAAGCTGAAG-----GACGGCGGCCTACTACTG 522
 Db 159 AsnGlyIleLysAlaAsnPheLysIleArgHisAsnIleGluAspGlyGlyValGlnLeu 178
 Qy 523 GTGAGGTTCT-----AAGTCCATCTACATGCCAAGAGCCCGTGCAGCTGCCCGGCTAC 576
 Db 179 AlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn 198
 Qy 577 TACTACGTGGACACCAAGCTGGACATCACC-----TCCCACACAGGAGGACTACACC 627
 Db 199 HisTyrLeuHisThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMet 218
 Qy 628 ATCGTGGAGCAGTACGAGCGCACCGAGGGCGGCCCGCCACCAC 666
 Db 219 ValLeuLeuGluPheValThrAlaAlaGlyIleThrHis 231

RESULT 10

US-09-023-946B-28
 ; Sequence 28 Application US/09023946B
 ; Patent No. 6670449

GENERAL INFORMATION:

APPLICANT: GERO MIESENBOCK, ET AL.
 TITLE OF INVENTION: HYBRID MOLECULES AND
 THEIR USE FOR OPTICALLY DETECTING CHANGES IN

CELLULAR MICROENVIRONMENTS

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:


```
/ APPLICATION NUMBER: US/09/023,946B
/ FILING DATE: 13-Feb-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/036,805
/ FILING DATE: 14 FEBRUARY 1997
/ APPLICATION NUMBER: 60/038,179
/ FILING DATE: 13 FEBRUARY 1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: KENNETH H. SONNENFELD
/ REGISTRATION NUMBER: 33,285
/ REFERENCE/DOCKET NUMBER: 2955-4004US2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 758-4800
/ TELEFAX: (212) 751-6849
/ TELEX: 421792
/ INFORMATION FOR SEQ ID NO: 28:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 238
/ TYPE: AMINO ACID
/ STRANDEDNESS: UNKNOWN
/ TOPOLOGY: UNKNOWN
/ SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-023-946B-28

Alignment Scores:
Pred. No.: 2,17e-12 Length: 238
Score: 228.50 Matches: 60
Percent Similarity: 47.21% Conservative: 50
Best Local Similarity: 25.75% Mismatches: 110
Query Match: 18.47% Indels: 13
DB: 4 Gaps: 6

US-10-081-864-14 (1-678) x US-09-023-946B-28 (1-238)
QY 1 ATGGCTCTCCGAGAACGTCATCACGAGTTCATCGGTTCAAGTGGCGATGAGGGC 60
Db 1 MetSerLysGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGCGCCCTACGAGGCG 120
Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGluGlyAspAlaThr-TyrGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 180
Db 41 LysLeuThrLeuLysPheIleCysThrThr--GlyLysLeuProValProThr 59
QY 181 CTGTCCCGCCAGTTCAGATCGGCTCAAGGTTCAGTGAAGCACCCTCCGACATCCC 240
Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
QY 241 -----GACTACAAAGAGCTCTCTCCCGAGGCTTCAAGTGGAGCGGTGATGAAC 294
Db 80 ArgHisAspPheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 99
QY 295 TTCAGAGACGGCGGTGGCGACCGTGAACCCAGGACTCTCTCCCTGAGGACGGCTGCTC 354
Db 100 PheLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ATCTACAAAGTGAAGTTCATCGGCTGAAGTTCATCGGCTGAAGTTCATCGGCTGA 414
Db 120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspLysAlaIleLeuGlyHis 139
QY 415 AAGACCATCCCAAGGCCCTCGAGCTGAAGTTCATCGGCTGAAGTTCATCGGCTGAAG 474
Db 140 Lys--LeuGluTyrAsnTyrAsnAspHisGlnValTyrIleMetAlaAspLysGlnLys 158
QY 475 GGGAGATCCCAAGGCCCTCGAGCTGAAGTTCATCGGCTGAAGTTCATCGGCTGAAG 522
Db 159 AsnGlyIleLysValAsnPheLysIleArgHisAsnIleGluAspGlyGlyValGlnLeu 178
QY 523 GTGAGATC-----AAGTCCATCTACATGGCCCAAGAACCCCTGAGCTGCGCGCTAC 576
Db 179 AlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn 198

US-10-081-864-14 (1-678) x US-08-818-604-32 (1-238)
QY 1 ATGGCTCTCCGAGAACGTCATCACGAGTTCATCGGTTCAAGTGGCGATGAGGGC 60
Db 1 MetSerLysGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGCGCCCTACGAGGCG 120
Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGluGlyAspAlaThr-TyrGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 180
Db 41 LysLeuThrLeuLysPheIleCysThrThr--GlyLysLeuProValProThr 59
QY 181 CTGTCCCGCCAGTTCAGATCGGCTCAAGGTTCAGTGAAGCACCCTCCGACATCCC 240
Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
QY 241 -----GACTACAAAGAGCTCTCTCCCGAGGCTTCAAGTGGAGCGGTGATGAAC 294
Db 80 GlnHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 99
QY 295 TTCAGAGACGGCGGTGGCGACCGTGAACCCAGGACTCTCTCCCTGAGGACGGCTGCTC 354
Db 100 TyrLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ATCTACAAAGTGAAGTTCATCGGCTGAAGTTCATCGGCTGAAGTTCATCGGCTGA 414
Db 179 AlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn 198

US-10-081-864-14 (1-678) x US-08-818-604-32 (1-238)
QY 1 ATGGCTCTCCGAGAACGTCATCACGAGTTCATCGGTTCAAGTGGCGATGAGGGC 60
Db 1 MetSerLysGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGCGCCCTACGAGGCG 120
Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGluGlyAspAlaThr-TyrGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 180
Db 41 LysLeuThrLeuLysPheIleCysThrThr--GlyLysLeuProValProThr 59
QY 181 CTGTCCCGCCAGTTCAGATCGGCTCAAGGTTCAGTGAAGCACCCTCCGACATCCC 240
Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
QY 241 -----GACTACAAAGAGCTCTCTCCCGAGGCTTCAAGTGGAGCGGTGATGAAC 294
Db 80 GlnHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 99
QY 295 TTCAGAGACGGCGGTGGCGACCGTGAACCCAGGACTCTCTCCCTGAGGACGGCTGCTC 354
Db 100 TyrLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ATCTACAAAGTGAAGTTCATCGGCTGAAGTTCATCGGCTGAAGTTCATCGGCTGA 414
Db 179 AlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn 198
```

Db 120 ValAsnArgIleGluLeuLysGlyLeuAspPheLysGluAspGlyAsnIleLeuGlyHis 139
 QY 415 AAGACCATGGCTGGAGGCGCTCCACGAGCGCTGTAC-----CCCCGC 459
 Db 140 Lys---MetGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysProLys 158
 QY 460 GACGCGGTCTGAAGGGCGAGATCCACAGCGCCCTGAAGCTGAAGGACGCGCGCCACTAC 519
 Db 159 AsnGlyIle---LysValAsnPheLysIleArgHisAsnIleLysAspGlySerValGln 177
 QY 520 CTGTGGAGTTC-----AAGTCCATCTACATGCGCCCAAGAGCGCGCTGACGCTGCCCGGC 573
 Db 178 LeuAlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAsp 197
 QY 574 TACTACTACGTGGACATACGAGCGCACCGAGGCGCGCCACAC 666
 Db 198 AsnHisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHis 217
 Db 218 MetIleLeuGluPheValThrAlaAlaGlyIleThrHis 231

RESULT 12

US-08-819-612-22
 ; Sequence 22, Application US/08819612D
 ; Patent No. 6172188
 ; GENERAL INFORMATION:
 ; APPLICANT: Thastrup, et al.
 ; TITLE OF INVENTION: No. 6172188el Fluorescent Proteins
 ; FILE REFERENCE: No. 6172188el Fluorescent Proteins
 ; CURRENT APPLICATION NUMBER: US/08/819.612D
 ; CURRENT FILING DATE: 1997-03-17
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 22
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Aequorea victoria
 US-08-819-612-22

Alignment Scores:
 Pred. No.: 3,27e-12 Length: 238
 Score: 226.50 Matches: 60
 Percent Similarity: 47.44% Conservative: 51
 Best Local Similarity: 25.64% Mismatches: 108
 Query Match: 18.31% Indels: 15
 DB: 3 Gaps: 7

US-10-081-864-14 (1-678) x US-08-819-612-22 (1-238)

QY 1 ATGGCTCTCCCGAGAACGTATCATCCGAGTTCATGGCTTCAAGTGGCATGGAGGC 60
 Db 1 MetSerLysGlyGluGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
 QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGAGGCGCGCCCTACGAGGC 120
 Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGlyGluGlyAspAlaThrGly 40
 QY 121 CACACACCGTGAAGTGAAGGTGACCAAGGCGCGCCCTCGCTTCCGCTGGGACATC 180
 Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThr 59
 QY 181 CTGTCCCGCCAGTTCACGATCGGCTCCAGGTGACGAGCACCACCGCGCATCCCC 240
 Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
 QY 241 -----GACTACAGAAGCTGTCTTCCCGAGGCTTCAAGTGGAGCGCGTGAAC 294
 Db 80 GlnHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 99
 QY 295 TTCGAGGACCGCGCGTGGCGACCGTACCCAGGACTCTCCCTCGGAGCGGCTCTTC 354
 Db 100 TyrLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119

QY 355 ATCTACAAGGTGAAGTTCATCGGCGTGAACCTCCCTCCGAGCGCGCGCTGATGAGAG 414
 Db 120 ValAsnArgIleGluLeuLysGlyLeuAspPheLysGluAspGlyAsnIleLeuGlyHis 139
 QY 415 AAGACCATGGCTGGAGGCGCTCCACGAGCGCTGTAC-----CCCCGC 459
 Db 140 Lys---MetGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysProLys 158
 QY 460 GACGCGGTCTGAAGGGCGAGATCCACAGCGCCCTGAAGCTGAAGGACGCGCGCCACTAC 519
 Db 159 AsnGlyIle---LysValAsnPheLysIleArgHisAsnIleLysAspGlySerValGln 177
 QY 520 CTGTGGAGTTC-----AAGTCCATCTACATGCGCCCAAGAGCGCGCTGACGCTGCCCGGC 573
 Db 178 LeuAlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAsp 197
 QY 574 TACTACTACGTGGACACCAAGCTGGACATCACC-----TCCACACGAGACTAC 624
 Db 198 AsnHisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHis 217
 QY 625 ACCATCGTGGAGCATGACGAGCGCACCGAGGCGCGCCACAC 666
 Db 218 MetIleLeuGluPheValThrAlaAlaGlyIleThrHis 231

RESULT 13

US-09-316-919-3
 ; Sequence 3, Application US/09316919
 ; Patent No. 6469154
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Baird, Geoffrey
 ; TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
 ; FILE REFERENCE: 07257/073001
 ; CURRENT APPLICATION NUMBER: US/09/316.919
 ; CURRENT FILING DATE: 1999-05-21
 ; NUMBER OF SEQ ID NOS: 63
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Aequorea victoria
 US-09-316-919-3

Alignment Scores:
 Pred. No.: 3,27e-12 Length: 238
 Score: 226.50 Matches: 60
 Percent Similarity: 47.44% Conservative: 51
 Best Local Similarity: 25.64% Mismatches: 108
 Query Match: 18.31% Indels: 15
 DB: 4 Gaps: 7

US-10-081-864-14 (1-678) x US-09-316-919-3 (1-238)

QY 1 ATGGCTCTCCCGAGAACGTATCATCCGAGTTCATGGCTTCAAGTGGCATGGAGGC 60
 Db 1 MetSerLysGlyGluGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20
 QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGAGGCGCGCCCTACGAGGC 120
 Db 21 AspValAsnGlyGlnLysPheSerValSerGlyGlyGluGlyAspAlaThrTyrGly 40
 QY 121 CACACACCGTGAAGTGAAGGTGACCAAGGCGCGCCCTCGCTTCCGCTGGGACATC 180
 Db 41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThr 59
 QY 181 CTGTCCCGCCAGTTCACGATCGGCTCCAGGTGACGAGCACCACCGCGCATCCCC 240
 Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
 QY 241 -----GACTACAGAAGCTGTCTTCCCGAGGCTTCAAGTGGAGCGCGTGAAC 294
 Db 80 GlnHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 99

```

QY 295 TTGAGGACGGCGGTGGCGACCGTGACCCAGGACTCCTCCCTCGACGAGCGGTGCTTC 354
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
100 TyrLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ACTACAAAGTGAAGTTCATCGCGGTGAAGTTCCTCCCTCGACGCGCCCGGTGATGACAG 414
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
120 ValAsnArgIleLeuLysGlyLeuAspPheLysGluAspGlyAsnIleLeuGlyHis 139
QY 415 AAGACCATGGGTGGGAGCGCTCCACCGAGCGCTGTAC-----CCCGCG 459
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
140 Lys---MetGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysProLys 158
QY 460 GACGGGTGCTGAAGGCGAGATCCCAAGGCGCTGAAGTGAAGGCGGCGGCGGACTAC 519
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
159 AsnGlyIle---LysValAsnPheLysIleArgHisAsnIleLysAspGlySerValGln 177
QY 520 CTGGTGGAGTTC-----AAGTCCATCTACATGGCCAGAGCGCGTGCAGTGCCTGGC 573
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 LeuAlaAspHisTyrGlnGlnAsnThrProLleGlyAspGlyProValLeuLeuProAsp 197
QY 574 TACTACTAGTGACACCAAGTGGACATCACC-----TCCACAAACGAGACTAC 624
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
198 AsnHisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHis 217
QY 625 ACCATCGTGAGCAGTACGAGCGCACCGAGGCGCGCCACCAC 666
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
218 MetIleLeuGluPheValThrAlaAlaGlyIleThrHis 231

RESULT 14
US-09-346-946-32
; Sequence 32, Application US/09346946
; Patent No. 6566083
; GENERAL INFORMATION:
; APPLICANT: Thastrup, Ole
; APPLICANT: Tullin, Soren
; APPLICANT: Poulsen, Lars
; APPLICANT: Bjorn, Sara
; TITLE OF INVENTION: A Method of Detecting Biologically
; FILE REFERENCE: 4301.204-US
; CURRENT APPLICATION NUMBER: US/09/346,946
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: US/08/818,604
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 0110/95
; PRIOR FILING DATE: 1995-01-31
; PRIOR APPLICATION NUMBER: 0982/95
; PRIOR FILING DATE: 1995-09-07
; PRIOR APPLICATION NUMBER: PCT/DK96/00052
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequora victoria
US-09-346-946-32

Alignment Scores:
Pred. No.: 3,27e-12 Length: 238
Score: 226.50 Matches: 60
Percent Similarity: 47.44% Conservative: 51
Best Local Similarity: 25.64% Mismatches: 108
Query Match: 18.31% Indels: 15
DB: Gaps: 7

US-10-081-864-14 (1-678) x US-09-346-946-32 (1-238)
QY 1 ATGGCTCTCCGAGAACGTATCACCGAGTTCAGCTTCAGGTGCGCATGAGGCG 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MetSerLysGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAspGly 20

```

```

QY 61 ACCGTGAACCGCCACGAGTTCGAGATCGAGGGCGAGGGCGCGCCCTAGGAGGC 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
21 AspValAsnGlyGlnLysPheSerValSerGluGluGlyAspAlaThrTyrGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGCACCAAGGGGGCGCCCTGCGCTTGGCTGGACATC 180
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
41 LysLeuThrLeuLysPheIleCysThrThr---GlyLysLeuProValProThr 59
QY 181 CTGTCCCGCCAGTTCAGTACGGTCCAGGTGATCGTGAAGCACCCCGCGGACATCCC 240
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
QY 241 -----GACTACAGAAGCTGCTCTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAAC 294
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
80 GlnHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluAspThrIlePhe 99
QY 295 TTCGAGGACGGCGCGTGGCGCACCGTACCCAGGACTCTCCCTCGAGGACGGGTGCTTC 354
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
100 TyrLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ATCTACAAAGTGAAGTTCATCGCGGTGAAGTTCCTCCCGAGGCGCGGTGATGACAG 414
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 139
QY 415 AAGACCATGGGTGGGAGCGCTCCACCGAGCGCTGTAC-----CCCGCG 459
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
140 Lys---MetGluTyrAsnTyrAsnSerHisAsnValTyrIleMetAlaAspLysProLys 158
QY 460 GACGGGTGCTGAAGGCGGAGATCCCAAGGCGCTGAAGTGAAGGCGGCGGCGGACTAC 519
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
159 AsnGlyIle---LysValAsnPheLysIleArgHisAsnIleLysAspGlySerValGln 177
QY 520 CTGGTGGAGTTC-----AAGTCCATCTACATGGCCAGAGCGCGTGCAGTGCCTGGC 573
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 LeuAlaAspHisTyrGlnGlnAsnThrProLleGlyAspGlyProValLeuLeuProAsp 197
QY 574 TACTACTAGTGACACCAAGTGGACATCACC-----TCCACAAAGGAGGACTAC 624
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
198 AsnHisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHis 217
QY 625 ACCATCGTGAGCAGTACGAGCGCACCGAGGCGCGCCACCAC 666
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
218 MetIleLeuGluPheValThrAlaAlaGlyIleThrHis 231

RESULT 15
US-09-023-946B-22
; Sequence 22, Application US/09023946B
; Patent No. 6670449
; GENERAL INFORMATION:
; APPLICANT: GERO MIESENBOCK, ET AL.
; TITLE OF INVENTION: HYBRID MOLECULES AND
; THEIR USE FOR OPTICALLY DETECTING CHANGES IN
; CELLULAR MICROENVIRONMENTS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,946B
; FILING DATE: 13-Feb-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/036,805
; FILING DATE: 14 FEBRUARY 1997
; APPLICATION NUMBER: 60/038,179

```

```
; FILING DATE: 13 FEBRUARY 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 2955-4004US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421752
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
; SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-09-023-946B-22
```

```
Alignment Scores:
Pred. No.: 3.27e-12 Length: 238
Score: 226.50 Matches: 60
Percent Similarity: 46.78% Conservative: 49
Best Local Similarity: 25.75% Mismatches: 111
Query Match: 18.31% Indels: 13
DB: 4 Gaps: 6
```

US-10-081-864-14 (1-678) x US-09-023-946B-22 (1-238)

```
QY 1 ATGCGCTCTCCGAGACGTCATCCAGGATTATCGCTTCAAGTGGCGATGGAGGC 60
Db 1 MetSerLysGlyGluLeuPheThrGlyValValProLeuValGluLeuAspGly 20
QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGAGGCGAGGCGGCCCTACGAGGGC 120
Db 21 AspValAsnGlyHisLysPheSerValSerGlyGluGluGlyAspAlaThrTyrGly 40
QY 121 CACAACACCGTGAAGTTCAGAGTGCACAGGGCGGCCCTCGCTTGGCTGGGACATC 180
Db 41 LysLeuThrLeuLysPheIleCysThrThr--GlyLysLeuProValProTyrProThr 59
QY 181 CTGTCCCCCGAGTTCAGTCCAGGTCCAGGTGAGTGAAGCAGCCCGCGACATCCCC 240
Db 60 LeuValThrThrPheSerTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 79
QY 241 -----GACTACAAAGAGTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGAAC 294
Db 80 ArgHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 99
QY 295 TTCGAGGCGCGCGTGGCGACCGTGCACCGAGCTCTCCCTGCGAGGCGGCTCTC 354
Db 100 PheLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 119
QY 355 ATCTACAAGTGAAGTTCATCGCGGTGAAGTCTCCCTCCGACGCGCCCGTGAAG 414
Db 120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 139
QY 415 AAGACCATGGCTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGACGCGCTGTAAG 474
Db 140 Lys---LeuGluTyrAsnTyrAsnAspHisGlnValTyrIleMetAlaAspLysGlnLys 158
QY 475 GCGGAGATCCACAGGCGCTGAAGCTGAAG-----GACGCGCGCCACTACCTG 522
Db 159 AsnGlyIleLysAlaAsnPheLysIleArgHisAsnIleGluAspGlyValGlnLeu 178
QY 523 GTGAGTTC-----AAGTCCATCTACATGGCCAGAGCCCGTGCAGCTGCCCGCTAC 576
Db 179 AlaAspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn 198
QY 577 TACTACGTGGACACCAAGCTGGACATCAC-----TCCCAACAGGAGGACTACACC 627
Db 199 HistyrLeuPheThrThrSerThrLeuSerLysAspProAsnGlnLysArgAspHisMet 218
QY 628 ATCGTGAGCAGTACGAGCGACCGAGGCGCGCCACCCAC 666
```

```
Db 219 ValLeuLeuGluPheValThrAlaAlaGlyIleThrHis 231
Search completed: July 29, 2004, 14:38:33
Job time : 17.9516 secs
```

OM nucleic - protein search, using frame_plus_n2p model
Run on: July 29, 2004, 14:36:21 / Search time 43.0787 Seconds
(without alignments)
9873.883 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 1237
Sequence: 1 atggcctctctcgagaacgt.....gccaccacctgttctctgttaa 678

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delopt 6.0, Delext 7.0

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 2582470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO spooll_p/US10081864/runat_29072004_150508_25599/app_query.fasta_1.1742
-DB=Published Applications AA -QWTF=fastan -SUFFIX=rabp -MINMATCH=0.1
-LOOPCL=0 -LOPEXI=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USRR=US10081864 @CGN 1.1.21 @runat_29072004_150508_25599
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Length DB ID Description
No. Score Match

1	1214	98.1	225	14	US-10-315-920-6	Sequence 6, Appli
2	1204	97.3	225	15	US-10-442-148A-7	Sequence 7, Appli
3	1204	97.3	225	15	US-10-442-148A-8	Sequence 8, Appli
4	1199	96.9	225	14	US-10-315-920-4	Sequence 4, Appli
5	1191	96.3	225	9	US-09-999-745-67	Sequence 67, Appli
6	1191	96.3	225	10	US-09-866-538-12	Sequence 12, Appli
7	1191	96.3	225	10	US-09-794-308-12	Sequence 12, Appli
8	1191	96.3	225	10	US-09-865-291-12	Sequence 12, Appli
9	1191	96.3	225	12	US-10-132-067-4	Sequence 12, Appli
10	1191	96.3	225	13	US-10-006-922-12	Sequence 8, Appli
11	1191	96.3	225	14	US-10-081-864-8	Sequence 1, Appli
12	1191	96.3	225	14	US-10-121-258-1	Sequence 2, Appli
13	1191	96.3	225	14	US-10-315-920-2	Sequence 56, Appli
14	1191	96.3	225	15	US-10-370-570-56	Sequence 32, Appli
15	1191	96.3	225	15	US-10-406-618-32	Sequence 13, Appli
16	1191	96.3	225	16	US-10-433-640-13	Sequence 1, Appli
17	1191	96.3	487	15	US-10-343-977-1	Sequence 2, Appli
18	1191	96.3	506	15	US-10-343-977-2	Sequence 3, Appli
19	1191	96.3	547	15	US-10-343-977-3	Sequence 2, Appli
20	1187	96.0	240	14	US-10-152-298-2	Sequence 2, Appli
21	1187	96.0	240	16	US-10-739-656-2	Sequence 2, Appli
22	1187	96.0	545	14	US-10-214-932-52	Sequence 52, Appli
23	1187	96.0	548	14	US-10-214-932-76	Sequence 76, Appli
24	1184	95.7	225	14	US-10-121-258-20	Sequence 20, Appli
25	1181	95.5	236	16	US-10-314-936-2	Sequence 2, Appli
26	1181	95.5	236	16	US-10-314-936-4	Sequence 4, Appli
27	1172.5	94.8	226	16	US-10-724-178-12	Sequence 12, Appli
28	1156	93.5	225	14	US-10-121-258-4	Sequence 4, Appli
29	1149	92.9	225	14	US-10-121-258-24	Sequence 41, Appli
30	1130	91.4	225	16	US-10-423-688A-41	Sequence 6, Appli
31	1124	90.9	225	14	US-10-121-258-6	Sequence 44, Appli
32	1104	89.2	225	13	US-10-006-922-44	Sequence 12, Appli
33	1104	89.2	225	14	US-10-081-864-12	Sequence 18, Appli
34	1070.5	86.5	230	13	US-10-006-922-18	Sequence 40, Appli
35	1070.5	86.5	230	14	US-10-161-403-40	Sequence 46, Appli
36	1023	82.7	205	13	US-10-006-922-46	Sequence 8, Appli
37	1021	82.5	225	14	US-10-121-258-8	Sequence 16, Appli
38	1021	82.5	225	16	US-10-724-178-16	Sequence 1041, Ap
39	846	58.4	183	16	US-10-724-178-1041	Sequence 1023, Ap
40	832	67.3	187	16	US-10-724-178-1023	Sequence 1037, Ap
41	793	64.1	168	16	US-10-724-178-1037	Sequence 10, Appli
42	762	61.6	227	13	US-10-006-922-10	Sequence 1033, Ap
43	762	59.5	152	16	US-10-724-178-1033	Sequence 5, Appli
44	726.5	58.7	232	14	US-10-133-973-5	Sequence 64, Appli
45	726.5	58.7	232	15	US-10-370-570-64	

ALIGNMENTS

RESULT 1
US-10-315-920-6
; Sequence 6, Application US/10315920
; Publication No. US20030175609A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.

US-10-315-920-6

Alignment Scores:

Pred. No.: 7,38e-84 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.14% Indels: 0
DB: 14 Gaps: 0

US-10-081-864-14 (1-678) x US-10-315-920-6 (1-225)

```
QY 1 ATGGCTCTCCGAGACGTCATCAGGCTTCATGCGCTTCAAGTGGCGATGGAGGC 60
DB 1 MetAlaSerSerGluAenValIleThrGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGCGGAGGGCCGCCCTTACGAGGC 120
DB 21 ThrValAenGlyHisGluPheGluIleGluGlyGluGlyA-gProTyrGluGly 40
QY 121 CACACACCGTGAAGTGAAGGTGACCAAGGGGGGGCCCTTCCCTTCCGCTGGGACATC 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCCCAGTTCACGAGTCCAGGTGACGAGTGTACGTGAAGCACCCCGCGGACATCCCC 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAAGAGTGTCTCTCCCGAGGGTTCAGTGGAGCGCGTGAATGAACTTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
DB 101 AspGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 120
QY 361 AAGTGAAGTTCATCGCGGTGAAGTTCCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAACTTCGAG 420
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGCCAAGAGCGGCTGCGAGTCCCGGCTACTACTACGTGAGCACCAGCTGGAC 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
QY 601 ATCACTCCCAACAGGAGTACACCATCGTGGAGCAGTACAGCGCGGCGGCGGCGGCGG 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTCTCTG 675
DB 221 HisHisLeuPheLeu 225
```

RESULT 2

US-10-442-148A-7

; Sequence 7, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
; APPLICANT: IWAKURA, MASAHIRO
; APPLICANT: HIROTA, KIYONORI
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
; TITLE OF INVENTION: PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
; FILE REFERENCE: 04583.0103-00000
; CURRENT APPLICATION NUMBER: US/10/442,148A
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: JP 2002-148950

; PRIOR FILING DATE: 2002-05-23

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 7

; LENGTH: 225

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-7

Alignment Scores:

Pred. No.: 4,24e-83 Length: 225
Score: 1204.00 Matches: 223
Percent Similarity: 99.11% Conservatives: 0
Best Local Similarity: 99.11% Mismatches: 2
Query Match: 97.33% Indels: 0
DB: 15 Gaps: 0

US-10-081-864-14 (1-678) x US-10-442-148A-7 (1-225)

```
QY 1 ATGGCTCTCCGAGACGTCATCAGGCTTCATGCGCTTCAAGTGGCGATGGAGGC 60
DB 1 MetAlaSerSerGluAenValIleThrGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGCGGAGGGCCGCCCTTACGAGGC 120
DB 21 ThrValAenGlyHisGluPheGluIleGluGlyGluGlyA-gProTyrGluGly 40
QY 121 CACACACCGTGAAGTTCAGGCTGACCAAGGGGGGGCCCTTCCCTTCCGCTGGGACATC 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCCCAGTTCACGAGTCCAGGTGACGAGTGTACGTGAAGCACCCCGCGGACATCCCC 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAAGAGTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAACTTCGAG 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
DB 101 AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTGAAGTTCATCGCGGTGAAGTTCCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAACTTCGAG 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCTGGGAGGCGCTCCACGAGCGCTGTACCCCGCGGAGCGCGTGTGAAGGGCGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGGCGCTTGAAGTGAAGTTCCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAACTTCGAG 540
DB 161 ThrHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGCCAAGAGCGGCTGCGAGTCCCGGCTACTACTACGTGAGCACCAGCTGGAC 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspThrLysLeuAsp 200
QY 601 ATCACTCCCAACAGGAGTACACCATCGTGGAGCAGTACAGCGCGGCGGCGGCGGCGG 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTCTCTG 675
DB 221 HisHisLeuPheLeu 225
```

RESULT 3

US-10-442-148A-8

; Sequence 8, Application US/10442148A
; Publication No. US20040014242A1

```

1  GENERAL INFORMATION:
2  APPLICANT: IWAKURA, MASASHIRO
3  APPLICANT: HIROTA, KIYONORI
4  TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
5  TITLE OF INVENTION: PROCESS FOR ARRANGING AND IMMOBILIZING PROTEIN USING THE SAME
6  FILE REFERENCE: 04583.0103-00090
7  CURRENT APPLICATION NUMBER: US/10/442,148A
8  CURRENT FILING DATE: 2003-05-21
9  PRIOR APPLICATION NUMBER: JP 2002-148950
10 PRIOR FILING DATE: 2002-05-23
11 NUMBER OF SEQ ID NOS: 12
12 SOFTWARE: PatentIn version 3.2
13 SEQ ID NO 8
14 LENGTH: 239
15 TYPE: PRT
16 ORGANISM: Artificial Sequence
17 FEATURE:
18 OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
19 US-10-442,148A-8

```

Alignment Scores:		
Pred. No.:	4.28e-83	Length: 239
Score:	1204.00	Matches: 223
Percent Similarity:	99.11%	Conservative: 0
Best Local Similarity:	99.11%	Mismatches: 2
Query Match:	97.33%	Indels: 0
DB:	15	Gaps: 0
US-10-081-864-14 (1-678) x US-10-442-1488-8 (1-239)		

US-10-081-864-14 (1-678) x US-10-442-148A-8 (1-239)

QY	1	ATGGCTCTCTCCGAGAACTGTCATACCGAGTTCATCGCTTCAGGTGGCGCATCGAGGCG	60
Db	1	MetAlaSerSerGluAsnValIleThrGluPheMetArgPheLysValArgMetGluGly	20
QY	61	ACCGTGAACGGCCACGAGTTCAGATCCAGGGCGAGGGCGAGGGCCGCCCTTACGAGGCG	120
Db	21	ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly	40
QY	121	CACACACCGTGAAGTTCAGGTGACCAAGGGCGGCCCTCGCTCCCTCCCTGGGACATC	180
Db	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle	60
QY	181	CTGTCCCCCGAGTTCACGTACGGTCCCAAGGTGTACGTGAAGCACCCCGCGACATCCCG	240
Db	61	LeuSerProGlnPheGlnTyr-GlySerLysValTyrValLysHisProAlaAspIlePro	80
QY	241	GACTACAGAAGCTGCTCTCCCGAGGGCTTCAAGTGGAGCGCGTGTAACTTCGAG	300
Db	81	AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
QY	301	GACGGCGCGTGGCGACCGTGACCGAGACTCTCTCCCTCGAGCAGCGCTCTTCATCTAC	360
Db	101	AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
QY	361	AAGGTGAAGTTCATCGGGTGAACCTCCCTCCGACGGCCCGTGTAGCAGAGAAGAC	420
Db	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
QY	421	ATGGGCTGGAGGCTCTCACCGAGGCGCTGTACCCCGCGACCGCGTGTGAAGGGCGAG	480
Db	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
QY	481	ATCCACAAGCGCCCTCAAGCTCAAGACCGCGCGCCACTACTCTGTGAGATTCAAGTCCATC	540
Db	161	ThrHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
QY	541	TACATGGCCCAAGAAGCCGTGCGTCCCGGCTACTACTACGTGGACACCACGCTGGAC	600
Db	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspAlaLysLeuAsp	200
QY	601	ATCACCTCCCAACAGGAGACTACACCATCGTGAGCAGTACGAGCGCACCGAGGCGCCG	660
Db	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220

```

QY      661 CACRACCTGTCTCTG 675
      |||||
Db      221 HHisLeuPheLeu 225

RESULT 4
US-10-315-920-4
; Sequence 4, Application US/10315920
; Publication NO. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCES: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-4

```

Alignment Scores:		
Pred. No.:	1.02e-82	225
Score:	1199.00	222
Percent Similarity:	98.11%	1
Best Local Similarity:	98.67%	2
Query Match:	96.93%	Mismatches:
DB:	14	Indels:
		Gaps:
		0

US-10-081-864-14 (1-678) x US-10-315-920-4 (1-225)

Qy	1	ATGGCTCTCTCGAGACGTCATCACCAGATTCTATGCGCTTCAAGGTGGCATGGAGGCG	60
Db	1	Meta:rgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly	20
Qy	61	ACCGTGAAAGGGCCACGAGTTCGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGC	120
Db	21	ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly	40
Qy	121	CACACACCGTGAAATTGAGGTACCAAGGGGGCCCCCTGGCTTGGCTGGGCATC	180
Db	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle	60
Qy	181	CTGTGCCCCCAGTTCACGTACGGCTCCAAGGTGTAGTGAAGCACCCCCCGGCACATCCCC	240
Db	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro	80
Qy	241	GACTACAAAGAGCTCTCTCCCCAGGGGTTCAAGTGGGAGGGCGCTGATGAATTCGAG	300
Db	81	AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Qy	301	GACGGGGCGTGGGACCGTGACCCAGGATCTCTCCCTCGAGACGGCTGCTTCATCTAC	360
Db	101	AspGlyGlyValAlaThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
Qy	361	AAGGTGAAGTTTCATCGGGGTCAATCTCCCTCCGACGGCCCCGTGATGCAGAGAAGACC	420
Db	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Qy	421	ATGGGCTGGAGGCTCCACCGACGGCTGTACCCCCCGACGGCGTGTGAAGGGCGAG	480
Db	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArcAspGlyValLeuLysGlyGlu	160

```
QY 481 ATCCACAGGCCCTGAAGCTGAAGACGGCGGCACCTACTCGTGTGAGTTCAAGTCCATC 540
Db 161 ILeHisAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGCCAGAGCCCGTGCAGCTCCCGGCTACTACTACCTGGACACCAAGCTGGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspThrLysLeuAsp 200
QY 601 ATCACTCCACACAGGAGCTACACCATCGTGCAGCAGTACGAGCGCCACCGAGGCGCGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCTCG 675
Db 221 HisHisLeuPheLeu 225
RESULT 5
US-09-999-745-67
; Sequence 67, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-999-745-67
Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 9 Gaps: 0
US-10-081-864-14 (1-678) x US-09-999-745-67 (1-225)
QY 1 ATGCGCTCTCCGAGAACGTATACCGAGTTATCGCTTCAAGTGGCGCATGGAGGC 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPheAlaTyrAspIle 60
QY 181 CTGTCCCCCAGTGTCTCTCCCGAGGCTTCAAGTGGAGCGCGGCGGCGGCGGCGGCGG 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAAGAGCTGTCTCTCCCGAGGCTTCAAGTGGAGCGCGGCGGCGGCGGCGGCG 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 100
QY 301 GACGGCGGCGTGGCGACCGTGAACCGAGTCTCTCCCGAGGCGGCGGCGGCGGCGGCGG 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTGAAGTTTCATCGGCGGTGAACCTTCCCTCCGACGGCGGCGGCGGCGGCGGCGG 420
```

```
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCTCGGAGGCGCTCCACCGAGCGCTGTACCCCGCGGCGGCGGCGGCGGCGGCGG 480
Db 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGGCCCTGAAGCTGAAGACGGCGGCACCTACTCGTGTGAGTTCAAGTCCATC 540
Db 161 ILeHisAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGCCAGAGCCCGTGCAGCTCCCGGCTACTACTACCTGGACACCAAGCTGGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspThrLysLeuAsp 200
QY 601 ATCACTCCACACAGGAGCTACACCATCGTGCAGCAGTACGAGCGCCACCGAGGCGCGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCTCG 675
Db 221 HisHisLeuPheLeu 225
RESULT 6
US-09-866-538-12
; Sequence 12, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-866-538-12
Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 10 Gaps: 0
US-10-081-864-14 (1-678) x US-09-866-538-12 (1-225)
QY 1 ATGCGCTCTCCGAGAACGTATACCGAGTTATCGCTTCAAGTGGCGCATGGAGGC 60
Db 1 MetArgSerSerLysAsnValIleLysGlnPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPheAlaTyrAspIle 60
QY 181 CTGTCCCCCAGTGTCTCTCCCGAGGCTTCAAGTGGAGCGCGGCGGCGGCGGCGGCGG 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAAGAGCTGTCTCTCCCGAGGCTTCAAGTGGAGCGCGGCGGCGGCGGCGGCG 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 100
```


Qy	301	GACGGCGCGTGGGACCCGGTGACCCAGGACTCTCTCCCTGCGAGGACGGTGCTTCATCTAC	360
Db	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIeIytr	120
Qy	361	RAGGTGAAGTTCATCGGCGGTGAACCTCCCTCCGACGGCCGCTGATCAGAGAAGACC	420
Db	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValIleGlnLysLysThr	140
Qy	421	ATGGGCTGGGAGGCTCCACCGAGGCGCTGTACCCCGCGCAGCGGCGTGTCTGAAGGGCGAG	480
Db	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
Qy	481	ATCCACAGGGCGCTGAAGCTGAAGACGGCGCCACTACCTGGTGGAGTTCAGTCCCATC	540
Db	161	IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
Qy	541	TACATGGCCAAAGACCCCGTGAGCTGCCCGGCTACTACTACGTGGACACCAAGCTGGAC	600
Db	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
Qy	601	ATCACCTCCCACACGAGGACTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCGCG	660
Db	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
Qy	661	CACGACCTGTTCTGTG	675
Db	221	HisHisLeuPheLeu	225

RESULT 7

```

US-09-794-308-12
; Sequence 12, Application US/09794308
; Publication No. US20030170311A1
; GENERAL INFORMATION
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: ZACHARIAS, David
; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REG15130
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-794-308-12

```

Alignment Scores:	4.12e-82	Length:	225
Pred. No.:		Matches:	220
Score:	1191.00	Conservative:	2
Percent Similarity:	98.67%	Mismatches:	3
Best Local Similarity:	97.78%	Indels:	0
Query Match:	96.28%	Gaps:	0
DB:	10		

US-10-081-864-14 (1-678) x US-09-794-308-12 (1-225)

QY	1	ATGGCCTCTCTCCGAGAACGTCATCACCGAGTTCATCGGCTTCAAGTGGCGCATGAGGGC	60
Db	1	MetArgSerSerIysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly	20
QY	61	ACCGTGAACGGCCACACAGTTCGAGATCGAGGGCGAGGGCGCGCCCTACGAGGGC	120
Db	21	ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly	40
QY	121	CACAAACACCGTGAAGTTGAAGGTGACCAAGGGGGGGCCCTCGCCCTTCGGCTGGGACATC	180
Db	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTyrAspIle	60
QY	181	CTCTCCCGCCAGTTCACAGTACGGCTCCAAAGGTACGTGAAGACACCCCGCGGCATCCCC	240

Db	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro	80
QY	241	GACTACAAGAAGCTGCTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACCTTCGAG	300
Db	81	AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
QY	301	GACGGGGCGTGGCGACCGTGACCCAGGACTCTCCCTGCAGAGCAGCGCTCTTCATCTAC	360
Db	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
QY	361	AAGGTGAAGTTTCATCGCGCTGAATCTCCCTCCGAGCGCCCGTGTGTCAGAGAAGAACC	420
Db	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
QY	421	ATCGGCTGGAGCCCTCCACCGAGCGCTGTACCCCGCGACGCGTGTGTAAGGCGGAG	480
Db	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyVal	160
QY	481	ATCCACAAGGCCCTGAAGTGTGAAGGACGGCGGCACCTACCTGTGTGGAGTTCAAGTCATC	540
Db	161	IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
QY	541	TACATGCCCAAGAAGCCCGTGGAGCTGCCCGGCTACTTACTACGTGGACACCAAGCTGCAC	600
Db	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
QY	601	ATCACTCTCCACAACAGGAGTACACCATCGTGGAGCAGTACAGAGCCACCGAGGGCGGC	660
Db	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
QY	661	CACCACCTGTTCTCTG	675
Db	221	HisHisLeuPheLeu	225

RESULT 8

```

US-09-865-291-12
/ Sequence 12, Application US/09855291
/ Publication No. US20030186229A1
/ GENERAL INFORMATION:
/ APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
/ APPLICANT: TSIEN, Roger
/ APPLICANT: TING, Alice
/ APPLICANT: ZHANG, Jin
/ TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
/ FILE REFERENCE: REGEN1550
/ CURRENT APPLICATION NUMBER: US/09/865,291
/ CURRENT FILING DATE: 2001-05-24
/ NUMBER OF SEQ ID NOS: 42
/ SOFTWARE: PatentIt version 3.0
/ SEQ ID NO 12
/ LENGTH: 223
/ TYPE: PRT
/ ORGANISM: Discosoma sp.

```

Alignment Scores:	4.12s-82	Length:	225
Pred. No.:	191.00	Matches:	220
Score:	98.67%	Conservative:	2
Percent Similarity:	97.78%	Mismatches:	3
Best Local Similarity:	96.28%	Indels:	0
Query Match:	10	Gaps:	0
DR:	10		

US-10-081-864-14 (1-678) X US-09-865-291-12 (1-225)

Qy	1	ATGCCCTCTCCGAGAACGTCATCACCGAGTTCATGCCCTTCAAGGTCGGCATCGAGGC	50
Db	1 <td>MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly</td> <td>20</td>	MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly	20
Qy	61	ACCGTGAACGCCCCACGAGTTCGAGATCCAGGCGCAGGCGCGGGCGGCCCTTACGAGGC	120
Db	21	ThrValLeuGluVHsGluPheGluIleGluGluGluGluGluVHsProtyrGluGly	40

QY 121 CACACACCGTGAAGTTGAAGTGAACCAAGCGCGCCCTGCGCTTCGCTGGACATC 180
 Db 41 HisAsnThrValLysLeuLysValThrLysGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCCTCCCGGAGTTCAGTACGGTCCCAAGGTGACGAGCAGCCCGCGCATCCCC 240
 Db 61 LeuSerProGlnPheGlnTyrGlySerLysValThrLysHisProAlaAspIlePro 80
 QY 241 GACTACAAGAGCTGCTCTCCCGAGGGCTTCAAGTGGAGCGCGTGAATCTCGAG 300
 Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGCGCGCGGCGGCGGCGTACCGAGCTCTCCCTCCCGAGCGCGCTCTCATCTAC 360
 Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGTGAAGTTCATCGCGTGAAGTTCAGGAGCGCGCGCTGATCGAAGAGAC 420
 Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 QY 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGAGCGCGTGAAGGCGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATCCACAGGCGCTGAAGTGAAGGAGCGCGCGCTACCTGTTGGAGTTCAAGTCCATC 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGGCGAGAGCGCGTGCAGCTCCCGGCTACTACTACCTGACACACAGCTGGAC 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 QY 601 ATCACTCCACACAGGAGGACTACACCATCGTGAGCAGTACGAGCGCACCGAGGCGCG 660
 Db 201 IleThrSerHisGlnAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACCACCTGTTCTCTG 675
 Db 221 HisHisLeuPheLeu 225

RESULT 9

US-10-132-067-4
 ; Sequence 4, Application US/10132067
 ; Publication No. US20030203355A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bradbury, Andrew
 ; APPLICANT: Zeytun, Ahmet
 ; APPLICANT: Waldo, Geoffrey
 ; APPLICANT: The Regents of the University of California
 ; TITLE OF INVENTION: Fluorobodies: Binding Ligands With Intrinsic
 ; TITLE OF INVENTION: Fluorescence
 ; FILE REFERENCE: 021362-00060005
 ; CURRENT APPLICATION NUMBER: US/10/132,067
 ; CURRENT FILING DATE: 2002-04-24
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 225
 ; TYPE: PRT
 ; ORGANISM: Discosoma sp.
 ; FEATURE:
 ; OTHER INFORMATION: red fluorescent protein (dsRED)
 US-10-132-067-4

Alignment Scores:
 Pred. No.: 4,12e-82 Length: 225
 Score: 1191.00 Matches: 220
 Percent Similarity: 98.67% Conservatives: 2
 Best Local Similarity: 97.78% Mismatches: 3
 Query Match: 96.28% Indels: 0
 DB: 12 Gaps: 0

US-10-081-864-14 (1-678) x US-10-132-067-4 (1-225)

QY 1 ATGGGCTCTCTCGAAGACGTGATCACCAGAGTTCATGCGCTTCAAGGTGCGGATCGAGGGC 60
 Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGCGGAGGCGGAGGCGGCGCTTACGAGGGC 120
 Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACACACCGTGAAGTTCAGGTCACCAAGGCGGCGGCGCTTCCCTTCCGCTGGAGATC 180
 Db 41 HisAsnThrValLysLeuLysValThrLysGlyProLeuProPheAlaTrpAspIle 60
 QY 181 CTGTCTCCCGGAGTTCAGTACGGCTCCAAAGGTGACGTGAAGCACCAGCGCGGACATCCC 240
 Db 61 LeuSerProGlnPheGlnTyrGlySerLysValThrLysHisProAlaAspIlePro 80
 QY 241 GACTACAAGAGCTGCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATCTCGAG 300
 Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGCGCGCTGGGAGCGCTGACCCGAGACTCTCCCTGCGAGGCGCTGCTTCTCATCTAC 360
 Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGTGAAGTTCATCGCGCTGAAGTTCCTCCCTCCGAGCGCGCTGATCGAAGAGAC 420
 Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 QY 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGAGCGCGTGTGAAGGCGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATCCACAGGCGCTGAAGTGAAGGAGCGCGCGCTACTACTGTTGGAGTTCAAGTCCATC 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGGCCAAAGACCGCGTGCAGCTCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 QY 601 ATCACTCCACACAGGAGGACTACACCATCGTGAGCAGTACGAGCGCACCGAGGCGCG 660
 Db 201 IleThrSerHisGlnAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACCACCTGTTCTCTG 675
 Db 221 HisHisLeuPheLeu 225

RESULT 10

US-10-006-922-12
 ; Sequence 12, Application US/10006922
 ; Publication No. US20020197676A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lukyanov, Sergey A
 ; APPLICANT: Pradkov, Arcady F.
 ; APPLICANT: Labas, Yulii A.
 ; APPLICANT: Matz, Mikhail V.
 ; APPLICANT: Tersikh, Alexey
 ; TITLE OF INVENTION: No. US20020197676A1 Chromophores/Fluorophores and
 ; FILE REFERENCE: CLON-035CIP
 ; CURRENT APPLICATION NUMBER: US/10/006,922
 ; CURRENT FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 09/120,330
 ; PRIOR FILING DATE: 1998-12-11
 ; PRIOR APPLICATION NUMBER: 09/457,898
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/458,144
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/458,477
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/457,556

Appl.


```
QY 601 ATCACTCCCAACAGAGGACTACACATCGTGAGCAGTACAGCGCACCGAGGGCGGC 660
Db 201 llethrSerHisnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACTGTTCTCTG 675
Db 221 HisHisLeuPheLeu 225
RESULT 12
US-10-121-258-1
; Sequence 1, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: U083-1CP2CF1
; CURRENT APPLICATION NUMBER: US/10/121,258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(225)
; OTHER INFORMATION: wild-type DsRed
US-10-121-258-1
Alignment Scores:
Pred. No.: 4.12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 14 Gaps: 0
US-10-081-864-14 (1-678) x US-10-121-258-1 (1-225)
QY 1 ATGCGCTCTCCGAGACGTCATCAGGATTCATCGGCTTCAAGTGCGCATGGAGGC 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGCCACAGTTCAGATCGAGCGGAGGCGGCGGCGCCCTACGAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGACCAAGGCGGCGGCGGCTTCGCTTCCGCTGGGACATC 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCGCCAGTCCAGTACCGCTCCAGGTGTAGTGAGCACCCCGCCGACATCCCC 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAGAGTTCATCGCGTGAAGTTCCTCCCTCCGAGCGGCGGCTGATGATGAGGAGGAC 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GAGGCGGCTGGGACCGGTGACCGAGGACTCTCCCTCGAGGACCGGCTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTTCATCGCGTGAAGTTCCTCCCTCCGAGCGGCGGCTGATGATGAGGAGGAC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
```

```
QY 421 ATGGGCTGGAGGCTCCACCGAGCGGCTGTACCCCGGAGCGGCTGCTCAAGGGCGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCCAAGGCGCTGAGCTGAAGCAGCGCGGCCACTACCTGTTGGAGTTCAAGTCCATC 540
Db 161 lleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCAAAGAGCCCGTGCAGCTGCCCGCTTACTACTACTACTACTACTACTGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACTCCCAACAGAGGACTACACATCGTGAGCAGTACAGCGCACCGAGGGCGGC 660
Db 201 llethrSerHisnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACTGTTCTCTG 675
Db 221 HisHisLeuPheLeu 225
RESULT 13
US-10-315-920-2
; Sequence 2, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Tetsikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-315-920-2
Alignment Scores:
Pred. No.: 4.12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 14 Gaps: 0
US-10-081-864-14 (1-678) x US-10-315-920-2 (1-225)
QY 1 ATGCGCTCTCCGAGACGTCATCAGGATTCATCGGCTTCAAGTGCGCATGGAGGC 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGCCACAGTTCAGATCGAGCGGAGGCGGCGGCGCCCTACGAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTGAAGTGACCAAGGCGGCGGCGGCTTCGCTTCCGCTGGGACATC 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCGCCAGTTCAGTACCGCTCCAGGTGTAGTGAGCACCCCGCCGACATCCCC 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAGAGTTCATCGCGTGAAGTTCCTCCCTCCGAGCGGCGGCTGATGATGAGGAGGAC 300
```

Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCGTGGCGACCGTGCACCGAGGACTCTCCCTGCGAGGCGGCTGCTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTGAAGTTCATCGGCGTGAAGTTCCTCCCTCGAGCGGCGGCTGCTTCATCTAC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGAGCGCTCCACCGAGCGCTGTACCCCGCGCGCGGCTGCTGAAGGGCGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGCCCTGAAGCTGAAGGACGCGCGGCGGCTGCTGCTGAGTTCAGTCCATC 540
Db 161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCCAAGAGCCGCTGAGCTGCCCGGCTACTACTGCTGCACACCAAGCTGGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACTCCCAACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCTCG 675
Db 221 HisHisLeuPheLeu 225

RESULT 14
US-10-370-570-56
; Sequence 56, Application US/10370570
; Publication No. US20030219717A1
; GENERAL INFORMATION:
; APPLICANT: DAHL, Soren Weis et al.
; TITLE OF INVENTION: FLUOROPHORE COMPLEMENTATION PRODUCTS
; FILE REFERENCE: 3759-0130P
; CURRENT APPLICATION NUMBER: US/10/370,570
; CURRENT FILING DATE: 2003-02-24
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 56
; TYPE: PRT
; ORGANISM: Diatoms sp.
US-10-370-570-56
Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 15 Gaps: 0
US-10-081-864-14 (1-678) x US-10-370-570-56 (1-225)
QY 1 ATGGCTCTCCGAGAGCGTCAACGAGTTCATCGCTTCAAGTGGCGGCTGATGAAGTTCGAG 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGCGGCGGCGGCGGCTGCTTCATCTAC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTCGAGGAGGACCAAGGGCGGCGGCTGCTTCGCTGGGACATC 180
Db 41 HisAsnThrValLysLysLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCGCCCGAGTTCAGTACGCTCCAGGTGTAGTGAAGCACCOCGCGCATCCCG 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

QY 241 GACTACAGAAGCTGTCTTCCCGAGGCGCTTCAAGTGGGAGCGGCTGATGAAGTTCGAG 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCGTGGCGACCGTGCACCGAGGACTCTCCCTGCGAGGCGGCTGCTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTGAAGTTCATCGGCGTGAAGTTCCTCCCTCGAGCGGCGGCTGCTTCATCTAC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGAGCGCTCCACCGAGCGCTGTACCCCGCGCGCGGCTGCTGAAGGGCGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGCCCTGAAGCTGAAGGACGCGCGGCGGCTGCTGCTGAGTTCAGTCCATC 540
Db 161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCCAAGAGCCGCTGAGCTGCCCGGCTACTACTGCTGCACACCAAGCTGGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACTCCCAACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCTCG 675
Db 221 HisHisLeuPheLeu 225

RESULT 15
US-10-406-618-32
; Sequence 32, Application US/10406618
; Publication No. US20030219814A1
; GENERAL INFORMATION:
; APPLICANT: Wan, David Chi-Cheong
; APPLICANT: Ip, Denis Tsz-Ming
; APPLICANT: The Chinese University of Hong Kong
; TITLE OF INVENTION: No. US20030219814A1el Fluorescent Proteins
; FILE REFERENCE: 016285-34-1
; CURRENT APPLICATION NUMBER: US/10/406,618
; CURRENT FILING DATE: 2003-04-02
; PRIOR APPLICATION NUMBER: US 60/387,968
; PRIOR FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: US 60/370,598
; PRIOR FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 32
; TYPE: PRT
; LENGTH: 225
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Spontaneously
; OTHER INFORMATION: Fluorescent protein DsRed.
US-10-406-618-32
Alignment Scores:
Pred. No.: 4,12e-82 Length: 225
Score: 1191.00 Matches: 220
Percent Similarity: 98.67% Conservative: 2
Best Local Similarity: 97.78% Mismatches: 3
Query Match: 96.28% Indels: 0
DB: 15 Gaps: 0
US-10-081-864-14 (1-678) x US-10-406-618-32 (1-225)
QY 1 ATGGCTCTCCGAGAGCGTCAACGAGTTCATCGCTTCAAGTGGCGGCTGATGAAGTTCGAG 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

QY 241 GACTACAGAAGCTGTCTTCCCGAGGCGCTTCAAGTGGGAGCGGCTGATGAAGTTCGAG 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGGCGTGGCGACCGTGCACCGAGGACTCTCCCTGCGAGGCGGCTGCTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTGAAGTTCATCGGCGTGAAGTTCCTCCCTCGAGCGGCGGCTGCTTCATCTAC 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGAGCGCTCCACCGAGCGCTGTACCCCGCGCGCGGCTGCTGAAGGGCGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAGCCCTGAAGCTGAAGGACGCGCGGCGGCTGCTGCTGAGTTCAGTCCATC 540
Db 161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCCAAGAGCCGCTGAGCTGCCCGGCTACTACTGCTGCACACCAAGCTGGAC 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACTCCCAACAGGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTTCTCG 675
Db 221 HisHisLeuPheLeu 225

QY 61 ACCGTGAACGGCCAGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGC 120
Db |||||||
21 ThrValAsnGlyHisGluPheGluileGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAACACCGTGAAGTTGAAGTCAACAAAGGGGGCCCTCCCTTCGCTCGGACATC 180
Db |||||||
41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 CTGTCCCCCAGTTCCTCCAGTACGGCTCCAAGTGTACGTGAAGCACCCCGCCGACATCCCC 240
Db |||||||
61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTACAAAGAGCTGCTCTCCCGAGGGCTTCAAGTGGAGCGCGTGTGATGAATTCGAG 300
Db |||||||
81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGCGCGTGGCGACCGCTGACCCAGGACTCTCTCCCTGCAGGACGCGCTGCTTCATCTAC 360
Db |||||||
101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTGAAGTTCATCGCGCTGAAGTCCCTCCGACGGCGCGCTGATGCGAGAGAGACC 420
Db |||||||
121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGACGCGCTGTGAAGGGCGAG 480
Db |||||||
141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATCCACAAGCCCTGAAGCTGAAGACGGCGGCGCCACTACCTGTGGAGTTCAGTCCATC 540
Db |||||||
161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCCAAGAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
Db |||||||
181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
QY 601 ATCACCTCCACACGAGGACTACCATCGTGGAGCAGTACGAGCGCACCGAGGCCCGC 660
Db |||||||
201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCACCTGTCTCTG 675
Db |||||||
221 HisHisLeuPheLeu 225

Search completed: July 29, 2004, 14:54:49
Job time : 46.0787 secs

OM nucleic - protein search, using frame_plus_n2p model
Run on: July 29, 2004, 14:28:05 ; Search time 14.6859 Seconds
(without alignments)
8881.684 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 1237
Sequence: 1 atggcctctccgagacgt.....gccaccacctgtctctgtaa 678

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2.1/USPTO.spool_p/US10081864/runat 29072004 150506 25492/app query.fasta_1.1742
-DB=PIR78 -QPMI=fastan -SUFFIX=pr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864 @CNG 1.1 53 @runat 29072004 150506 25492 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DST TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 78: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225.5	18.2	238	1 JQ1514	green-fluorescent
2	205.5	16.6	580	2 T43481	probable mucin DKF
3	203	16.4	3020	2 A43932	mucin 2 precursor,
4	186	15.0	1106	2 JQ4005	hypothetical i19.5
c 5	182	14.6	1791	2 T02345	hypothetical prote
6	175.5	14.2	279	2 S53363	mucin SAC (clone J
c 7	175	14.0	660	1 QQB83	BHLFI protein - hu
8	174	14.1	317	2 S55316	mucin (clone PGM-2
9	171	13.8	377	2 A48018	mucin 7 precursor,
10	170.5	13.8	528	2 I47141	gastric mucin (clo
11	170.5	13.8	1367	1 S48478	glucan 1,4-alpha-g
12	168.5	13.6	335	2 T05722	licheninase (EC 3.
13	165.5	13.4	1161	2 S57180	probable membrane
14	164	13.3	383	2 T46707	proteophosphoglyca

15	163	13.2	507	2 T44768	antifreeze glycope
c 16	163	13.0	749	2 A45294	Balbani ring 2.1
17	162.5	13.1	283	2 S13383	hydroxyproline-ric
18	160	12.9	839	2 T04859	extensin homolog F
19	159.5	12.9	267	2 S08314	cell wall glycopro
20	159.5	12.9	529	2 T45134	hypothetical prote
21	159	12.9	464	2 S22697	extensin - Volvox
22	157	12.7	350	2 S22456	hydroxyproline-ric
23	157	12.7	494	2 G84348	hypothetical prote
24	156	12.6	525	2 JN0059	hypothetical 57.4K
c 25	154.5	12.4	221	2 C34768	ORF2 protein - Orf
26	154.5	12.5	866	2 T45462	membrane glycoprot
c 27	153	12.2	414	2 C86301	arginine/serine-ri
28	153	12.4	770	2 T22808	hypothetical prote
29	152.5	12.3	164	2 I53641	mucin SAC - human
30	152	12.3	264	2 S29893	salivary glue prot
31	152	12.3	303	2 S28364	hydroxyproline-ric
32	152	12.3	867	2 T45463	membrane glycoprot
33	151.5	12.2	797	1 VGBEX1	glycoprotein X pre
34	150.5	12.2	222	2 H96711	hypothetical prote
35	150.5	12.2	477	2 S53362	mucin SAC (clone J
c 36	150.5	12.0	1520	2 T44231	hypothetical prote
37	150.5	12.2	3570	2 T45025	mucin MUC5B, trach
38	150	12.1	328	2 JQ0985	hydroxyproline-ric
39	150	12.1	707	2 A46302	PTB-associated spl
c 40	149.5	12.0	528	2 B75310	conserved hypotnet
41	148.5	12.0	1513	2 A54895	mucin 2, intestina
c 42	148	11.8	319	2 F75420	hypothetical prote
43	148	12.0	322	2 A53715	apomucin precursor
44	147.5	11.9	214	2 T10737	extensin-like cell
45	147.5	11.9	660	2 JN0057	chitinase (EC 3.2.

ALIGNMENTS

RESULT 1

JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C:Accession: J50692; JQ1514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JQ1514; MUID:92175527; PMID:11347277
A:Accession: J50692
A:Molecule type: DNA
A:Residues: 1-107,'S',109-238 <PRA1>
A:Cross-references: GB:M62654; NID:g155662; PIDN:AAA27722.1; PID:g155663
A:Accession: JQ1514
A:Molecule type: mRNA
A:Residues: 1-99,'F',101-140,'L',142-218,'V',220-238 <PRA2>
A:Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64,74-122,132-151,154-183,185-200 <PRA3>
R:Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24,'Q',26-156,'P',158-171,'K',173-238 <INO>
A:Cross-references: GB:I29345; NID:g606383; PIDN:AAA58246.1; PID:g606384
R:Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13,'V',15-24,'Q',26-44,'N',46-153,'G',155-156,'P',158-171,'K',173-227,'R',
A:Cross-references: EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PID:g634009
A:Experimental source: clone gfp1

R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A;Reference number: A43932; MUID:91358717; PMID:1885763
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g18863; PIDN:AAA59875.1; PID:g188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A;Reference number: A33532; MUID:89197956; PMID:2703501
A;Accession: E33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257; MUID:91086481; PMID:1985113
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner,
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t
A;Reference number: PQ0328; MUID:92198477; PMID:1550588
A;Accession: PQ0328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: PQ0329
A;Molecule type: protein
A;Residues: 2328-2342, 'K', 2344-2354 <XUG1>
C;Genetics:
A;Gene: GDB:MUC2
A;Cross-references: GDB:120203; OMIM:158370
A;Map position: lip15.5-lip15.5
C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
C;Keywords: glycoprotein; intestine; tandem repeat
F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

Alignment Scores:

Pred. No.:	Length:	Score:
7.62e-07	3020	
203.00	Matches:	74
37.90%	Conservative:	20
29.84%	Mismatches:	104
16.41%	Indels:	50
2	Gaps:	9

US-10-081-864-14 (1-678) x A43932 (1-3020)

Qy	5	CCTCCTCCGAGAACGTCAATCACCGAGTTTCATCGCGTTCACGGTGCCCATGGAGGCACCG	64
Db	1518	ProThrThrThrThrProSerProThrThrThrThrThr-----ThrProPro	1534
Qy	65	TGAACGCCACAGTTCAGATCGAGCGGCGAGGCGCGCCCTACGAGGCACCA	124
Db	1535	ProThrThrThrProSerProThrThrThrProleThrProThrSerThrThr	1554
Qy	125	ACA---CGTGAAGTTGAAGTGACAAAGGCGGCCCCCTCGCCTTGGGACATCC	181
Db	1555	ThrLeuPro-----ProThrThrThrProSerProThrThrThr	1569
Qy	182	TGTCCTCCCGCATTCAGTAGCGCTCCAAGGTAGTGAAGCACCCCGGACATCCC	241
Db	1570	ThrThrProProProThrThrThrProSerProThrThrThrThrProSerProPro	1589

QY 623 ACACCATGTTGGAGCAGTACGAGCGCCAGCGGCGCCACCACC 667
 Db ValProThrThrSerThrThrSerAlaSerThrThrThrThrThr 249
 RESULT 7
 QMBE3
 BHLF1 protein - human herpesvirus 4 (strain B95-8)
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
 C:Accession: A03742
 R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrall, B.G.
 Mol. Biol. Med. 1, 21-45, 1983
 A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
 A:Reference number: A93065; MUID:85035713; PMID:6092825
 A:Accession: A03742
 A:Molecule type: DNA
 A:Residues: 1-660 <BAN>
 R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
 Nature 310, 207-211, 1984
 A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
 A:Reference number: A03794; MUID:84270667; PMID:6087149
 A:Contents: annotation; protein coding region
 C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
 C:Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:
 Pred. No.: 6.57e-05 Length: 660
 Score: 175.00 Matches: 77
 Percent Similarity: 33.22% Conservative: 17
 Best Local Similarity: 27.21% Mismatches: 83
 Query Match: 14.00% Indels: 106
 DB: 1 Gaps: 16
 US-10-081-864-14 (1-678) x QMBE3 (1-660)

QY 673 GGACAGGTGTGGCGGC----- 656
 Db GlyProGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySer 218
 QY 655 -----CTCGGTGGCTCGTACGTCTCCAGGTGGTGTAGTCTCTCT 614
 Db GlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnProArgLeu 238
 QY 613 -----TGTGGAGGTGATGTCACAGT----- 593
 Db ProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArgSerGly 258
 QY 592 -----TGTGTCCAGTGTAGT----- 575
 Db AlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAenPro 278
 QY 574 -----ACCGGCGAGCTGCACGGCTCTTGGCCATGTAGATCG 536
 Db GlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHisPro----- 295
 QY 535 ACTTGAATCCACAGGTAGTCCCGCGCTCTTACAGTTCAGGGCCTTGTGATCTCGC 476
 Db ProAlaAlaProGlyAlaProGlyThrProAla-----AlaProGlyProGlyGly 327
 QY 415 TCTTCTCATCACGGGCGCTCGGAGGGGAAGTTCACGCCGATGAACCTTCACTGTGAGA 356
 Db GlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyProAla--- 346
 QY 355 TGAAGCAGCGCTCTCGAGGAGAGTCTGTGGGTACGGTCCGCCAGCGCGCTCTCGA 296
 Db AspProProAlaAlaAlaArgLeuProProGluArgGlnProArgLeuProGln 365
 QY 295 -----AGTTTCATCACCGCTCCCACTTGAAGCCCTCGGGAAGGACGAGCTTCTGT 245

Db 366 AspLeuAlaAlaAlaGlnArgCysProAlaGlyProProThrArgSerGlyAlaAla 385
 QY 244 AGTCGG-----GGATGTCGGCGGTGCTTCCACGTACACCTTGG 206
 Db 386 AlaGlnArgThrHisArgArgProProGlyCysProArgSerAla---ArgAsnPro--- 403
 QY 205 AGCCGTACTGGAACCTGGGGGACAGATGTCACAGGGGAGAGGGGCGGCCCTTGG 146
 Db 404 -----GlyCysPro-----ArgThrTrp 409
 QY 145 TCACCTTCACTTCACGGTGTGTGGCCCTCGTAGGGGCGCCCTCGCCCT----- 95
 Db 410 -----ArgArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaGly 425
 QY 94 ---CGCCCTCGATCTCGAACTCGTGGCGGTTCCAGGTGCCCT-----CCA 53
 Db 426 GlnArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrPro 445
 QY 52 TCGCCACCT 44
 Db 446 AlaAlaPro 448
 RESULT 8
 S55316
 mucin (clone PGM-2B) - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 17-Nov-2000
 C:Accession: S55316
 R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulos-Cladaras, M.; Specian, R.D.; LaMont, J.T.
 Biochem. J. 308, 89-96, 1995
 A:Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.
 A:Reference number: S55315; MUID:95275264; PMID:7755593
 A:Accession: S55316
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-317 <TUR>
 A:Cross-references: GB:U12768; NID:G915204; PIDN:AAC48525.1; PID:G915207
 C:Superfamily: pig submaxillary mucin

Alignment Scores:
 Pred. No.: 7.1e-05 Length: 317
 Score: 174.00 Matches: 81
 Percent Similarity: 45.23% Conservative: 28
 Best Local Similarity: 33.61% Mismatches: 92
 Query Match: 14.07% Indels: 40
 DB: 2 Gaps: 10
 US-10-081-864-14 (1-678) x S55316 (1-317)
 QY 8 CCTCCGAGAACGTCA-----CCGAGTTCA 34
 Db 51 ProProGluThrSerSerHisGlyAlaThrSerSerThrThrThrSerValGlnProSerSer 70
 QY 35 TGCCTTCAGTCCGATGAGGGCAGCGTGAACGCCACGATTCGATCGAGGGCG 94
 Db 71 SerSerSer-----AlaProThrThrSerAlaThrSerValGlnProSer 85
 QY 95 AGGCGAGGGCGCCCTACGAGGGCCACACCGTGAAGTTGAAGTTGACCAAGGGCG 154
 Db 86 SerSerGlySerAlaProThrThrSerAlaThrSerValGlnSerSerSerGlySer 105
 QY 155 GCCCCTGCTCCCTCGCTGGGACATCTGTCGCCCGCCAGTTCCAGTCCAGTCCCA 208
 Db 106 AlaProThrThrSer---AlaThrSerValGlnProSerSerSerSerSerProIle 124
 QY 209 AGGTGTACGTGAACACCCCGCGACATCC-----CCGACTACAAGAGCTGT 256
 Db 125 SerSerThrIleSerValGlnProSerSerSerSerSerAlaProThrThrSerAlaThr 144
 QY 257 CTTTCCCGAGGGCTTCAAGTGGAGCGCTGATGAACATTTCGAGGAGCGCGCGTGGCGA 316
 Db 145 SerValGlnSerSerSerGlySerAlaProThrThrSerAlaThrSerValGlnPro 164

```

QY 317 CCGTGACCCAGGACTCCCTCCCTGCAGGACGGCTGCTTCA---TCTACAAGGTGAAGTCA 373
DB 165 SerSerSerSerSerProProProSerSerSerSerSerSerSerSerSerSerSerSer 184
QY 374 TCGCGGTGAAGTCCCTCCGAGCCCGCGGTGATCCAGAGACCATGGCTGGGAGG 433
DB 185 SerAlaProThrThrSerAlaThrSerVal-----Gln 195
QY 434 CTTCCACCGAGCGCTGTACCCCGCGAGCGGGTCTGAGGGCGAGATCCACAAGGCC 493
DB 196 SerSerSerSerSerSerAlaProThrThrSer-----AlaThrSerValGlnPro 212
QY 494 TGAAGCTGAAGCGCGCGCCACTACTCTGGTGGAGTTCAAGTCCATCTACATGGCCAGA 553
DB 213 SerSerSerGlySerAlaProThrThrSerAlaThrSerValGlnSerSerSerSer 232
QY 554 ACCCGGTGAGCTGCGCGCTACTACTAGTGACACCAAGCTGGACATCACTCCCA 613
DB 233 SerProProSerSerThrThrSerValGlnThrSerSerSerSerSerSerProThr 252
QY 614 ACGAGGACTACA-----CCATCGTGGAGCAGTACGAGCGCACCGAGGCCGCCACC 664
DB 253 ThrSerThrThrSerValGlnProSer---SerSerGlySerAlaProThrThrSerAla 271
QY 665 ACC 667
DB 272 Thr 272

RESULT 9
A48018
mucin 7 precursor, salivary - human
N/Alternate names: mucin, MG2; mucin, MG2a-T1; mucin, MG2a-T2; mucin, MG2b-T2
C/Species: Homo sapiens (man)
C/Date: 16-Feb-1994 #sequence revision 18-Nov-1994 #text_change 07-May-1999
C/Accession: A48018; S29115; S29116; S29114
R/Bobek, L.A.; Tsai, H.; Biesbrock, A.R.; Levine, M.J.
J. Biol. Chem. 268, 20563-20569, 1993
A/Title: Molecular cloning, sequence, and specificity of expression of the gene encoding
A/Reference number: A48018; MUID:93388636; PMID:7690757
A/Accession: A48018
A/Molecule type: mRNA
A/Residues: 1-377 <BOB>
A/Cross-references: GB:113283
A/Experimental source: submandibular gland
A/Note: sequence extracted from NCBI backbone (NCBIN:137719, NCBI:P:137720)
R/Reddy, M.S.; Bobek, L.A.; Haraszthy, G.G.; Biesbrock, A.R.; Levine, M.J.
Biochem. J. 287, 639-643, 1992
A/Title: Structural features of the low-molecular-mass human salivary mucin.
A/Reference number: S29114; MUID:93075006; PMID:1445223
A/Accession: S29115
A/Molecule type: mRNA
A/Residues: 143-168 <RED>
A/Accession: S29116
A/Molecule type: protein
A/Residues: 'S', 71-79, 'N', 81-86, 'XX', 89, 'X', 91, 'P' <RED>
A/Accession: S29114
A/Molecule type: protein
A/Residues: 143-145, 'X', 147, 'XXX', 151-152, 'X', 154-158, 'X', 160-161, 'A', 163-164, 'XX', 167-171
C/Genetics:
A/Gene: GDB:MUC7
A/Cross-references: GDB:138799; OMIM:158375
A/Map position: 4q13-4q21
A/Keywords: glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-377/Product: mucin 7, salivary #status predicted <MAT>
F:97,128,135,146,312/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 0.000119 Length: 377
Score: 171.00 Matches: 75
Percent Similarity: 40.00% Conservative: 23
Best Local Similarity: 30.61% Mismatches: 88

```

Query Match: 13.82% Indels: 59
DB: 2 Gaps: 10

US-10-081-864-14 (1-678) x A48018 (1-377)

```

QY 1 ATGGCTCTCCGAGAACGTCATCAGGAGTTCATGGCTTCAAGGTGGCA-----52
DB 139 ILeSerSerArgGluAsnValAenThr-SerSerSerValAlaThrLeuAlaProValas 158
QY 53 -----TGGAGGGCACCGTGAAGCCGACCGAGTTCGAGATCGAG 90
DB 158 nSerProAlaProGlnAspThrThrAlaAlaProProThrProSerAlaThrThrProAl 178
QY 91 GGCAGGGCGAGCGCCCGCTACGAGGGCCACACACCGTGAAGTGAAGGTGACCAAG 150
DB 178 aProProSerSerSerAlaProProGluThrThrAla-----190
QY 151 GGGGGCGCCCTGCGCTCGGACATCTCTCCCGCCAGTTCAGTCCAGTCCGCGGC 210
DB 191 -AlaProProThrProSerAlaThrThrGlnAlaProProSerSerSerAlaProProGl 210
QY 211 GTGTACGTGAAGACCCCGCGGACATCCCGACATCCAGAGAGTGTCTTCCCGGAGGC 270
DB 210 uThrThr---AlaAlaProProThrProProAlaThr-----ThrProAlaProProSe 227
QY 271 TTCAAGTGGGCGCGGTGATGAATTCGAGAGCGCGCGTGGCGACCGTGACCCAGGAC 330
DB 227 rSerSerAlaPro-----ProGluThrThrAlaAla-----ProProTh 240
QY 331 TCCTCCCTCGAGGACGGCTGCT-----352
DB 240 rProSerAlaThrThrProAlaProLeuSerSerSerAlaProProGluThrThrAlaVa 260
QY 353 -----TCATCTACAAGTGAAGTTCATCGCGTGAATTCCTCCCTCCGAGGC 399
DB 260 lProProThrProSerAlaThrThrLeuAspProSerSerSerAlaSerAlaProProGluTh 280
QY 400 CCCGTGATGCAGAGAAGACCATCGGTGGAGCGCTCCACCGAGCGCTGATACCCCGC 459
DB 280 rThr-----AlaAlaProProThrProSerAlaThrThrProAl 293
QY 460 GACGGGTGTGAAGGGGAGATCCACAGCCCTGAGCTGAAGGACGGCGGCCTACTAC 519
DB 293 aPro-----ProSerSerProAlaProGlnGluThrThrAlaAlaProIleTh 309
QY 520 CTGGTGGAGTTCAGTTCATCATCGGCAAGAGCCCGTGCAGCTGCCCGGTACTACTAC 579
DB 309 rThrProAsnSerSerProThrThr-----LeuAlaProAspThrSerGluThrSerAl 327
QY 580 TACGTGCACACCAAGCTGGAGCATCCTCCCAACAGAGAGTACACCATCGTGGAGCAG 639
DB 327 aAlaProThrHisGlnThrThrThrSerValThrGlnThrThrThrThrThrLysGlnPr 347
QY 640 TACGAGCGCACCG 652
DB 347 oThrSerAlaPro 351

```

RESULT 10

I47141
gastric mucin (clone PGM-2A) - pig (fragment)
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000
C/Accession: I47141; S55315
R/Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Gastroenterology 106, 200, 1994
A/Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a novel
A/Reference number: I47141; MUID:94102478; PMID:7506218
A/Accession: I47141
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-528 <TUR>
A/Cross-references: EMBL:U10281; NID:G915205; PIDN:AAC48526.1; PID:G915208
R/Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.

184 ----- 184
QY 535 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 554
Db
QY 185 -----CCCCCAGTTCAGTACAGTCTCAGTGAAGACACCCCGCCGACATCCCG 241
Db
QY 555 ProThrProSerSerThrThrGluSerSer-----SerThrProValThrSerSer 572
Db
QY 242 ACTACAGA-----AGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGAAT 295
Db
QY 573 ThrThrGluSerSerSerAlaProValProThrProSerSerSerThr-----Thr 589
Db
QY 296 TCAGAGACGGCGGTGGCGACCGTGACCC-----AGGACTCCT 334
Db
QY 590 GluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerAla 609
Db
QY 335 CCTGCGACGAGCGCTCTCATCTACAAGTGAAGTTCATCGCGTGAATCTCCCTCG 394
Db
QY 610 ProAlaProThrProSerSerSerThrThrGluSerSerSerAla-----Pro 625
Db
QY 395 ACGGCCCCGTGATGACAGAGAAGACCATGGGTGGAGCGCTCCACCGAGCCCTGTACC 454
Db
QY 626 ValThrSerSerThrThrGluSerSerSerAlaProValProThrProSerSerThr 645
Db
QY 455 CCGCGACGGGTGTGAGGGCGAGATCCACAGGCCCTGAAGCTGAAGGACGGCGGC 514
Db
QY 646 ThrGluSer-----SerSerAlaProValProThrProSerSerSer 659
Db
QY 515 ACTACCTGTGTGAGTTCAGTTCATCTACATGCCCAAGAAGC-----CCGTGC 562
Db
QY 660 ThrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSer 679
Db
QY 563 AGTGTCCCGGTCTACTACTAGTGGACACCAAGTGGACATCACTCCCAACAGGACT 622
Db
QY 680 SerAlaProValThrSerSerThrThrGluSerSerSerAlaProValThrSerSerThr 699
Db
QY 623 -----ACACATCGTGGAGCAGTACAGCGCACCGAGG 655
Db
QY 700 ThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSer 719
Db
QY 656 GCCGCCACCACTGTTCCT 674
Db
QY 720 Ala-ProValProThrPro 725
Db

RESULT 12

T05722
Licheninase (EC 3.2.1.73) - barley
N;Alternate names: lichenase
C;Species: Hordeum vulgare (barley)
C;Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 08-Oct-1999
R;Slakeski, N.; Baulcombe, D.C.; Devos, K.M.; Ahluwalia, B.; Doan, D.N.P.; Fincher, G.B.
Mol. Gen. Genet. 224, 437-449, 1990
A;Title: Structure and tissue-specific regulation of genes encoding barley (1->3, 1->4)-
A;Reference number: S13734; MUID:91094782; PMID:2266947
A;Accession: T05722
A;Status: preliminary; translated from GH/EMBL/DBEJ
A;Molecule type: DNA
A;Residues: 1-335 <SLA>
A;Cross-references: EMBL:X56260; NID:g18991; PIDN:CAA39700.1; PID:g18992
A;Experimental source: cv. NK 1559
C;Genetics:
A;Map position: 1L
A;Introns: 25/2
C;Keywords: glycosidase; hydrolase

Alignment Scores:
Pred. No.: 0.000178 Length: 335
Score: 168.50 Matches: 77
Percent Similarity: 33.80% Conservative: 19
Best Local Similarity: 27.11% Mismatches: 89
Query Match: 13.62% Indels: 99
DB: 2 Caps: 12

RESULT 13

S57180
Probable membrane protein YJR151c - Yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J2223; serine/threonine-rich protein YJR151c
C;Species: Saccharomyces cerevisiae
C;Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 19-Apr-2002
C;Accession: S57180
R;Scarce, T.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57169

US-10-081-864-14 (1-678) x T05722 (1-335)

QY 18 CGTCATCACGAGTTCATCGCTTCAAGGTGCGCATCGAGGGCAC----- 62
Db
QY 34 ArgHisGluArgGlnSerAlaGlyGlyGluHisArgGlyGlnHisValGlnValGln 53
Db
QY 63 -----CGTGAACGGCCACGAGTTCGAGATCCAGGCGGAGG 98
Db
QY 54 ArgAspGlnLeuHisAlaAlaValArgSerArgProGlyGlyAlaAlaGlyGlyArg 73
Db
QY 99 CGAGGG-----CGCCCTACGAGG 119
Db
QY 74 HisGlyArgGluArgCysGlyArgAlaGlnArgAlaLeuGlnProArgArgGln 93
Db
QY 120 CCACAACACCGTGAAGTGAAGTGAAGGGGGCCCTGCTCCCTCGCTGGGACAT 179
Db
QY 94 SerArgSerGlyCysIleValGlyGluGlnHisPro-----GlyVal 108
Db
QY 180 CTTGTCCCCCAGTTCAGTACGGCTCAAGGTGTACGTGAAGCACCCCGGACATCCC 239
Db
QY 109 ProGlnGlyLeuLeuProValCysLeuArgGlyGlnArgGlyArgArgGlyHisPro 128
Db
QY 240 CGACTACAAGAAGCTGCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATTCGA 299
Db
QY 129 Glu-----ProCysProArgHisGluGluArgAlaGlyArgAlaGlyLeuArg 144
Db
QY 300 -----GGACGGCGGTGGGAC-----CGTGACCCGAGACTCCTC 335
Db
QY 145 ArgAlaGlyProHisGlnGlyAspHisValGlyValAlaGlyHisProGlyGlyValGln 164
Db
QY 336 CCTCAGACGGCTGCTTCTATCTACAAGTGAAGTTCATCGCGCTGAATCTCCCTCCGA 395
Db
QY 165 ProAlaValArgArgVal-----LeuHisArgArgGlyGlyArgValHis 179
Db
QY 396 CGGCCCCGTGATGCAGAGAAGACCATGGGTGGGAGCGCTTCACCGAGCGCTGTACC 455
Db
QY 180 GlyProArgGlyAla-----ValProCysArgPro 188
Db
QY 456 CCGGACGGCTGTGAGGGCGAGATCCACAGGCCCTGAAGTGAAGGACGGCGGCA 515
Db
QY 189 HisArgArgAlaHisGlyGln-----HisLeuPro 199
Db
QY 516 CTACTGTGT---GGAGTTCAGTCCATCTACATCGCCCAAGAGCCCGTGCAGCTGCCGG 572
Db
QY 200 ValProGlyLeuGlyLeuGlnProGluArgHisGlyHisGlyLeuArgAlaLeuHisArg 219
Db
QY 573 CTACTACTGTGACACACCAAGCTGGACATCCTCCCAACAGGAGACTACACCATCGT 632
Db
QY 220 LeuArgHisArgGlyProGlyArgLeuLeuArgValProGluProValA-GHisHisArg 239
Db
QY 633 GGAGCAGTA----- 641
Db
QY 240 GlyArgLeuLeuHisGlyHisGlyGlnAlaArgArgLeuGlnArgGluAlaGlyGlyVal 259
Db
QY 642 -----CGAGCGCACCGAGGGCGGCA----- 662
Db
QY 260 ArgGluArgValAlaValSerArgArgHisGlyAspProGlyGlnArgGlnAspLeu 279
Db
QY 663 CCACCTGTCTCT 674
Db
QY 280 GlnProValPro 283
Db

Search completed: July 29, 2004, 14:37:30
Job time : 23.6859 secs

US-10-081-864-14 (1-678) x T44768 (1-507)

[illegible]

Blank Sheet USPTO

OM nucleic - protein search, using frame_plus_n2p model
Run on: July 29, 2004, 14:24:20 ; Search time 9.79061 Seconds
(without alignments)
7211.716 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 1237
Sequence: 1 agggcctctcgagacgt.....gccaccacctgtctctgtaa 678

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cpn2_1/USPTO.spool.p/US10081864/runat 29072004 150505 25463/app query.fasta_1.1742
-DB=SwissProt 42 -QPMF=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10081864 @CN 1.1.22 @runat 29072004 150505 25463 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224.5	18.1	238	1 GFP_AEQVI	P42212 aequorea vi
2	203	16.4	5179	1 MUC2_HUMAN	Q02817 homo sapien
3	175	14.0	660	1 YHLI_EBV	P03181 Epstein-Barr
4	170.5	13.8	1367	1 AMYL_YEAST	P08640 saccharomyc
5	165.5	13.4	1161	1 DAN4_YEAST	P47179 saccharomyc
6	161	12.9	1033	1 IP2_STRCO	Q8CJG8 streptomyc
7	159.5	12.9	267	1 EXTN_MAIZE	P14318 zea mays (m
8	153.5	12.4	283	1 EXTN_SORBI	P24152 sorghum bic
9	151.5	12.2	797	1 VGLX_HSVB	P28968 equine hepi
10	150.5	12.2	5703	1 M5B_HUMAN	Q9HC84 homo sapien
11	150	12.1	707	1 SFPQ_HUMAN	P23246 homo sapien
12	150	12.0	1046	1 IP2_STRAW	Q82K53 streptomyc
13	148.5	12.0	1513	1 MUC2_RAT	Q62635 rattus norv
14	148	11.8	659	1 SFRG_HUMAN	Q8N2M8 homo sapien
15	146	11.8	1233	1 M5A_HUMAN	P98088 homo sapien
16	145.5	11.8	437	1 YGY3_HALSQ	F21561 haloferax s
17	145.5	11.8	573	1 C114_MOUSE	P19467 mus musculu
18	145.5	11.8	1336	1 W146_HUMAN	Q9C0J8 homo sapien

19	144.5	11.7	307	1 SGS3_DROME	P02840 drosophila
20	144.5	11.6	1461	1 IE18_PRVIF	P11675 pseudorabie
21	144	11.6	600	1 SP96_DICDI	P14328 dictyosteli
22	143.5	11.5	348	1 YT35_STRFR	P20186 streptomyc
23	143.5	11.5	497	1 VE2_HPV20	P50766 human papil
24	142	11.4	653	1 SFRG_MOUSE	Q8CFC7 mus musculu
25	142	11.4	825	1 ICPG_HSV2H	P28284 herpes simp
26	141.5	11.4	297	1 SGS4_DROME	Q00725 drosophila
27	141.5	11.4	555	1 GP1_CHLRE	Q09P66 chlamydomon
28	141	11.4	217	1 SGS3_DROS1	P13729 drosophila
29	140.5	11.2	514	1 VE2_HPV5B	P26545 xenopus papil
30	139.5	11.3	662	1 MUC1_XENLA	Q05049 xenopus lae
31	139.5	11.3	699	1 VGLG_HSV2H	Q63003 rattus norv
32	139.5	11.2	825	1 SE5_RAT	P21290 herpes simp
33	139	11.1	1446	1 IE18_PRVKA	P33479 pseudorabie
34	138	11.2	1509	1 GSRI_HUMAN	Q9NZM4 homo sapien
35	137.5	11.1	349	1 ENL2_ARATH	Q9T076 arabidopsis
36	137.5	11.1	1733	1 VNUA_PRVKA	P33485 pseudorabie
37	137	11.0	502	1 VE2_HPV25	P36787 human papil
38	137	11.0	1226	1 CDN1_HUMAN	Q81WY9 homo sapien
39	136	11.0	261	1 PRP2_MOUSE	P05142 mus musculu
40	136	10.9	514	1 VE2_HPV05	P06921 human papil
41	136	10.9	2424	1 CCAA_RABIT	P27884 oryctolagus
42	136	11.0	3149	1 TEGU_EBV	P0186 Epstein-Bar
43	135.5	10.8	431	1 UL61_HCMVA	P16818 human cytom
44	135	10.9	263	1 SGS3_DROYA	P13728 drosophila
45	133.5	10.7	512	1 IE63_HSV2H	P28276 herpes simp

ALIGNMENTS

RESULT 1	GFP_AEQVI	STANDARD;	PRT;	238 AA.
ID	GFP_AEQVI	STANDARD;		
AC	P42212; Q17104; Q27903;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cornier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent			
RT	protein."			
RL	Gene 111:229-233 (1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and			
RT	fluorescence characteristics of the recombinant protein."			
RL	FEBS Lett. 341:277-280 (1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97299832; PubMed=9154981;			
RA	Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;			
RT	"Enhanced expression in tobacco of the gene encoding green fluorescent			
RT	protein by modification of its codon usage."			
RL	Plant Mol. Biol. 33:989-999 (1997).			
RN	[4]			
RP	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea			
RT	green-fluorescent protein."			
RL	Biochemistry 32:1212-1218 (1993).			

Best Local Similarity: 25.32% Mismatches: 111
Query Match: 18.15% Indels: 13
DB: 1 Gaps: 6

US-10-081-864-14 (1-678) x GFP_AEQVI (1-238)

```
QY 1 ATGCGCTCTCCGAGACGTCATCCGAGTTCATCGCTTCAAGGTGGCATGAGAGGC 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MetSerlysglygluLeuLeuphethrglyvalProilleuvalgluLeuAspIy 20
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 ACCGTGAACGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCCTCAGAGGC 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21 AspValasnGlyHisLysPheSerValserGlyGluGlyAspAlathrTyrgly 40
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 CACACACCGTGAAGTGAAGGTGACCAAGGGGGGGCCCTCGCTTCGCTGGGACATC 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41 LysLeuthrLeuLysPheIleCysThrThr--GlyLysLeuProValProThr 59
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 CTGTCCCCCAGTTCACAGTTCAGGTTCACAGGTTCAGTGAAGCACCCGCCACATCCC 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60 LeuValThrThrPheSerTyrglyValGlnCysPheSerArgTy:ProAspHisMetLys 79
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 -----GACTACAAGAAGTGTCTCTCCGAGGGCTTCAAGTGGAGCGCGTGTGAAC 294
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 80 GlnHisAspPhePheLysSerAlaMetProgluglyTy:ValGlnGluArgThrIlePhe 99
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 295 TTCGAGGACGGCGGTGGCGACGTGACCCAGACGTCCTCCCTGTCAGGACGGCTGCTC 354
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 PheLysAspAspGlyAsnTy:LyseThrArgAlaGluValLysPheGluGlyAspThrLeu 119
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 355 ATCTACAAGGTGAAGTTCATCGCGGTGAAGTTCCTCCGACGGCCCCGTGATGAGAG 414
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 139
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 415 AGACATCGGCTGGGAGGCTCCACGAGCGGCTGTACCCCGACGGCGTGTCTGAAG 474
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140 Lys---LeuGluTy:AsnTy:AsnSerHisAsnValTy:ileMetAlaAspLysGlnLys 158
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 475 GCGGAGATCCACAGGCGCTCAAGCTGAAG-----GACGGCGGCCACATACCTG 522
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159 AsnGlyIleLysValAsnPhelyIleArgHisasnIleGluAspGlySerValGlnLeu 178
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 523 GTGAGATTC-----AAGTCATCATGTCGCAAGAGCCCGTCGAGTGTCCCGGTAC 576
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 AlaAspHisTy:GlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsn 198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 577 TACTATGGGACACCAAGCTGGACATACC-----TCCACAAACGAGACATACCC 627
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 199 HisTy:LeuSerThrGlnSerAlaLeuSerLysAspProasnGluLysargaspHisMet 218
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 628 ATCTGGAGCATGACGAGCGCACCGAGGGCGGCCACCAAC 666
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 ValLeuLeuGluPheValThrAlaAlaGlyIleThrHis 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 2

MUC2_HUMAN STANDARD; PRT; 5179 AA.

```
ID ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC AC Q02817; Q14878;
DT DT 01-JUN-1994 (Rel. 29, Created)
DT DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE Mucin 2 precursor (Intestinal mucin 2).
GN GN MUC2 OR SMUC.
OS OS Homo sapiens (Human).
OC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX OX NCBI_TaxID=9606;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC TISSUE=Intestine;
RX RX MEDLINE=94132002; PubMed=8300571;
RA RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
```

RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RN SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RP TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RA "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RN SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RP MEDLINE=91358717; PubMed=1885763;
RX Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RA "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism".
RL J. Clin. Invest. 88:1005-1013(1991).
CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and
CC other mucus membrane-containing organs. Thought to provide a
CC protective, lubricating barrier against particles and infectious
CC agents at mucosal surfaces.
CC -!- SUBUNIT: Multimeric.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -!- PM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY!).
CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and
CC varies among different alleles.
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC -!- SIMILARITY: Contains 2 VWFC domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L21998; AAB95295.1; -;
DR EMBL; M74027; AAA59875.1; -;
DR EMBL; M94131; AAA59163.1; -;
DR EMBL; M94132; AAA59164.1; -;
DR FIR; A49963; A43932.
DR Genew; HGNC:7512; MUC2.
DR MIM; 158370; -;
DR GO; GO:0005803; C:secretory vesicle; TAS.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_F.
DR Pfam; PF00007; Cys_knot; 1.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00214; VWC; 2.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01208; VWFC_1; 2.
DR PROSITE; PS0184; VWFC_2; 2.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 5179 MUCIN 2.
FT DOMAIN 1401 1747 APPROXIMATE REPEATS.

DB: 1 Gaps: 9

US-10-081-864-14 (1-678) x DAN4_YEAST (1-1161)

QY 17 AGCTCATCACGAGTTCATGGCTTCACAGTGGCGATGGAGGCGACCGTGAAGCGGCACG 76
D 143 ThrSerThrThrSerThrThrSerThr-----ThrProThrThrSerThr 157
QY 77 AGTTTCAGATCGAGGCGGAGGCGGAGGCGCGCCCTACGAGGGGCGACACACCGTGAAGT 136
D 158 ThrSer-----ThrThrProThrThrSerThrThr----- 167
QY 137 TGAAGGTGACCAAGGGCGCGCCCTCGCTCGCTGGGACATCTGTCGCCCGCAGTTC 196
D 168 -----SerThrThrProThrThrSerThrThrSerThrThrThrSerThr 183
QY 197 AGTACGGCTCAAGGTGTACGTGAAGCACCAGCGCGGACATCCCGAGTACA-----AGAAGC 253
D 184 ThrThrSerThrThrProThrThrSerThrThrSerThrThrProThrThrSerThrThr 203
QY 254 TGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAATTCGAGAGCGCGCGGTGG 313
D 204 SerThrThrProThrThrSerThrThrSerThrThrProThrThrSerThrThrSerThr 223
QY 314 CGACGTGACCCAGGACTCTCTCCCTCGAGGAGCGGTCTTCATCTACAAGGTGAAGTTCA 373
D 224 ThrProThrThrSerThrThrProThrThrSerThrThrSerThrThrSerGluThrSer 243
QY 374 TCGGGGTGAACCTTCCCTCCGACGCGCGCGCGGTGATGCAGAGAAGACACATGGGTGGGAGG 433
D 244 ThrLysSerThrThrProThrThrSer----- 252
QY 434 CTTCCACCGAGCGCTGTACCCCGCGAGCGGTGCTGAAGGGCGAGATCCACAAGGCC 493
D 253 -----SerThrSerThrThrProThrThr-----SerThrPro 264
QY 494 TGAAGCTGAAGGACGCGCGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAGA 553
D 265 -----ThrThrSerThrThr-----SerThrAlaProThrThr----- 275
QY 554 AGCCCGTGCAGTCCCGCGCTACTACTAGTGGACACCAAGCTGGACATCACTCCACCA 613
D 276 -----SerThrThrSerThrThrSerThrThrSerThrThrSerThrAlaProThr 292
QY 614 ACGAGGACTACACCATCTGGGAGGAGTACGAGCGGACCGAGCGCGCCACCACTGTTC 673
D 293 ThrSerThrThrSerThrThrPheSerThrSerSerAlaSerAlaSerSerValIleSer 312
RESULT 6
IF2_STRCO STANDARD; PRT; 1033 AA.
AC Q8CJ08;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR SC05706 OR SC9F2.10C OR SC5H4.30.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID:1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kiese R.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kiese R., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Batrell B.G., Parkhill J.,
RA

RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC !- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (by similarity).
CC !- SUBCELLULAR LOCATION: Cytoplasmic.
CC !- SIMILARITY: Belongs to the IF-2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL939124; CAD55362.1; --
CC PIR: T35989; T35989.
CC HAMAP: MF_00100; --, 1.
CC InterPro: IPR000637; AT_hook.
CC InterPro: IPR001851; Bac_inmem_transp.
CC InterPro: IPR004161; EFTU_D2.
CC InterPro: IPR000795; EF_GTPbind.
CC InterPro: IPR000178; IF2.
CC InterPro: IPR006847; IF2_N.
CC InterPro: IPR001806; Rag_transfrmng.
CC InterPro: IPR005225; Small_GTP.
CC Pfam: PF00009; GTP_EFTU; 1.
CC Pfam: PF03144; GTP_EFTU_D2; 1.
CC Pfam: PF04760; IF2_N_2.
CC PRINTS: PR00929; ATHOOK.
CC PRINTS: PR0315; ELONGATNFACT.
CC PRINTS: PR00449; RASTRNSFRMNG.
CC ProDom: PD186100; IF2; 1.
CC TIGRFAMS: TIGR00487; IF-2; 1.
CC TIGRFAMS: TIGR00231; small GTP; 1.
CC PROSITE: PS01176; IF2; FALSE NEG.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 529 681 G-DOMAIN.
FT NP_BIND 535 542 GTP (BY SIMILARITY).
FT NP_BIND 585 589 GTP (BY SIMILARITY).
FT NP_BIND 639 642 GTP (BY SIMILARITY).
SQ SEQUENCE 1033 AA; 105657 MW; 47DS81FB0072A045 CRC64;

Alignment Scores:
Pred. No.: 0.00588 Length: 1033
Score: 161.00 Matches: 65
Percent Similarity: 38.61% Conservative: 13
Best Local Similarity: 32.18% Mismatches: 78
Query Match: 12.88% Indels: 46
DE: 1 Gaps: 11

US-10-081-864-14 (1-678) x IF2_STRCO (1-1033)
QY 655 CCTCGTGGCGCTCGTACTGCTCCACGATGCTGTAAGTCTCTGTTGGGAGGTGATGTC 596
D 135 ProArgProAlaProAlaProAlaProGluProThrAlaPro-----Pro 148
QY 595 GCTTGGTGTCCAGTGTAGTAGTACCGCGGAGCTGCACGGGCTCTTGGCCATGTAGATGG 536
D 149 Ala---AlaProAlaProAlaProSerThrProAlaProAlaProSerGlyPro----- 164
QY 535 ACTTGAACCTCCACAGGTAGTGGCCCGCTCTCTCAGCTTTCAGGGCTTGTGGATCTCGC 476
D 165 -----LysProGly---GlyAlaAlaGProGlyAlaProLysProGlyAlaArg 180
QY 475 CTTTCA-----GCAGCCCTCCGGGGGTACAGGCGCTCGTGGAGGCGCTCCAGCCCA 421
D

```

181 ProSerGlyProGlyGlnAspArgGlyGlnGlnGlyGlnGlyArgPro----- 197
421 TGGTCTCTTCGATCATACGGGGCGCTCGAGGGAAGTTACACCGCATGAATTCACCT 362
198 -----GlyGlnArgProGlyAlaPro----- 205
361 TGTAGTAGAAGCAGCGCTCTCGAGGAGGAGTCTCGGTGACGGTCCGACCGCCGCGT 302
206 -----AlaGlnArgProGlyGlyArg-----ProGlyGlyProArgProGlyAsnAsn 221
301 CCTGAAAGTTCA-----TCAGCGCTCCCACTTGAAGCCCTCGGGGA 260
222 ProPheThrSerGlyGlyAsnAlaGlyMetAlaArgProSerAlaProArgProGlnGly 241
259 AGGACAGCTCTTGTAGTCGGGATGTCGGGGGGTCTTCACGTACACCTTGAGCCGT 200
242 GlyProArgProGlyGlyProGlyGlyAlaProGlyGlyGlyProArgProGlnGlyPro 261
199 ACTGGAAGTGGGGGACAGGATGTCACAGGCGAAGGCGAGGCGCGCCCTTGTCACCT 140
262 GlyGlyGlnGlyGlyProArgProGlnAlaProGlyGlyAsnArgProSerProGly 281
139 TCACTTCACGG-----TGTTGGGCTCTGT---AGGGGCGCCCTCGCCCT 95
282 SerMetProArgProGlnGlyGlyGlyAlaGlyProArgProGlyGlyProArgPro 301
94 CGCCCT 89
302 AsnPro 303

RESULT 7
EXTN MAIZE STANDARD; PRT; 267 AA.
AC P14918;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Extensin precursor (Proline-rich glycoprotein).
GN HRGP.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NGBI_taxid=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. W64A, and cv. E41;
RA Stiefel V., Perez-Grau L., Albericio F., Giralt E., Ruiz-Avila L.,
RA Ludevid M.D., Puigdomenech P.
RT "Molecular cloning of cDNAs encoding a putative cell wall protein from
RL Zea mays and immunological identification of related polypeptides.";
CC Plant Mol. Biol. 11:483-493(1988).
CC -!- FUNCTION: Structural component in primary cell wall.
CC -!- TISSUE SPECIFICITY: Mainly in the coleoptile node and root tip.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13499; CAA31854.1; -
DR EMBL; X13506; CAA31860.1; -
DR EMBL; M36912; AAA33455.1; -
DR EMBL; M36913; AAA33456.1; -
DR EMBL; M36914; AAA33457.1; -
DR PIR; S08314; S08314.

```

```

DR MaizeDB; 17152; -
DR InterPro; IPR003882; Pistill_extensin.
DR PRINTS; PRO1218; PSTILTEXTENSIN.
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW Hydroxylation.
FT SIGNAL 1 ?
FT CHAIN ? 267
FT DOMAIN 18 253
FT REPEAT 18 33
FT REPEAT 34 54
FT REPEAT 55 70
FT REPEAT 71 91
FT REPEAT 92 107
FT REPEAT 108 128
FT REPEAT 129 144
FT REPEAT 145 160
FT REPEAT 161 179
FT REPEAT 180 195
FT REPEAT 196 211
FT REPEAT 212 232
FT REPEAT 233 253
FT DOMAIN 261 265
FT VARIANT 245 245
FT VARIANT 261 261
SQ SEQUENCE 267 AA; 28349 MW; A6F406F4645FEBCB CRC64;

Alignment Scores:
Pred. No.: 0.00684 Length: 267
Score: 159.50 Matches: 75
Percent Similarity: 39.32% Conservative: 17
Best Local Similarity: 32.05% Mismatches: 94
Query Match: 12.89% Indels: 48
DB: 1 Gaps: 14

US-10-081-864-14 (1-678) x EXTN_MAIZE (1-267)
QY 5 CCTCTCCGAGAACGTATCACCAGTTCATCCCTTCAGAGTCGGATCGGATGAGGCGACCG 64
DB 34 ProProThrTyThrProSerProLysProAlaSerLysProProThrProLysPro 53
QY 65 TGAACGGCCACGAGTTCGAGATCGAGGCGGAGGCGGCGCCCTACGAGGCCACA 124
DB 54 ThrProProThrTyThrProSer-----ProLysProProThrProLysPro 69
QY 125 ACACCGTGAAGTTGAAGGTGACCAAGGGCGGCGCCCTCGCTTCCGCTGGGACATCTGT 184
DB 70 ThrPro-----ProThrTyThrProSerProLysProProAlaThrLys 84
QY 185 CCCCCAGT---TCCAGTAGCGCTCCAGGTGTACGTGAAGC-----ACCCCG 229
DB 85 ProProThrProLysProProThrTyThrProSerProLysProProThrPro 104
QY 230 ---CCGACATCCCGCATACAGAAGCTGCTTCCCGAGGGCTTCAAGTGGAGCGCG 286
DB 105 LysProThrProProThrTyThrProSerProLysProProAlaThrLys----- 121
QY 287 TGAAGAACTTCGAGGACGGCGGCTGGCGACCGGACCCAGG---ACTCCTCCCTGCAGG 343
DB 122 -----ProProThrProLysProProThrTyThr 132
QY 344 ACGGCTGCTTCATCTACAAGGTGAAGTTTCATCGCGTGAACCTCCCTCCGAGGCGCCCG 403
DB 133 ThrPro-----SerProLysProProThrProLysProProThrPro 146
QY 404 TGATGCGAGAGAACCATCGGCTGGGAGGCGCTCCACCGAGCGCTGTATCCCGCGGACG 463
DB 147 ThrTyThrProSerPro-----LysProProThrProLysProProThr 163
QY 464 CGGTGCTGAAGGGCGGAGATCCCAAGGCCCTGAAGCTGAAGCGGCGGCGGCGGCTACCTGG 523
DB 164 -----TyrThrProSerProLysProProThrHisProThrProLysProProThr 180
QY 524 TGGAGTTCAAGTCCATCTACATGCCAAGAGCCCGTGGAGCTCCCGGCGGTACTACTACG 583

```

```

Db 181 ProThrTyrThrProSer-----ProLysProProThrProLysProProThrProProThr 198
QY 584 TGACACCAAGTGGACATCACTCCACACAGGAGTACACCATCGGGAGCAGTACG 643
Db 199 TyrThrProSer---ProLysProProThrProLysProProThrProPro---ThrTyrThr 216
QY 644 AGGCACACGAGGCGCCCA-----CCACTGTTCCT 674
Db 217 ProSerProLys-ProProAlaThrLysProProThrPro 229

RESULT 8
EXTN_SORBI
ID EXTN_SORBI STANDARD; PRT; 283 AA.
AC P24152;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Extensin precursor (Proline-rich glycoprotein).
GN HRGP.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACAD clade; Panicoideae; Andropogoneae; Sorghum.
OX NCBI_TaxID=4559;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=91370882; PubMed=1893107;
RA Raz R., Cretin C., Puigdemonech P., Martinez-Izquierdo J.A.;
RT "The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum
vulgare.";
RL Plant Mol. Biol. 16:365-367(1991).
CC -!- FUNCTION: Structural component in primary cell wall.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
GLYCOSYLATED.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56010; CAA39485.1; -.
DR PIR; S13383;
DR InterPro; IPR003882; Pistil_extensin.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW Hydroxylation.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 283 EXTENSIN.
SQ SEQUENCE 283 AA; 29593 MW; 8D7FCD0DA8ED2D90 CRC64;

Alignment Scores:
Pred. No.: 0.0162 Length: 283
Score: 153.50 Matches: 67
Percent Similarity: 35.94% Conservative: 25
Best Local Similarity: 26.17% Mismatches: 76
Query Match: 12.41% Indels: 88
DB: 1 Gaps: 14

US-10-081-864-14 (1-678) x EXTN_SORBI (1-283)
QY 5 CTTCTCCGAGACGTGTCATCCAGTTCATCGCTTCAAGTGGCGATGGAGGCGCCG 64
Db 82 ProProProAlaThrProLysProProThr-----Pro 91
QY 65 TGAACGCCACG---AGTTTCAGATCGAGGCGGAGGCGGAGGCGCCCTACGAGGCG 121
Db 92 ProThrTyrThrProSerProLysProLysSerProValTyrProProProLysAla 111

```

```

QY 122 ACAACACCGTGAAGTGAAGGCGGCGCCCTGCTGCTTCCCTGGGACATCC 181
Db 112 SerThrPro-----ProThrTyrThrProSerProLysProProAlaThr 126
QY 182 TGTCTCCCGGAGTCCAGTACGCTCCCAAGGTGACGTGAAGCACCCCGGACATCCCG 241
Db 127 LysProProThrTyrProThr---ProLysProProAlaThrLysProProThrProPro 145
QY 242 ACTACAAAGAACTGCTCTCCCGAGGCTTCAAGTGGAGCGCGTGAATTCGAGG 301
Db 146 ValTyrThrProSerProLysProProValThrLys----- 157
QY 302 ACGGCGGCGTGGCGACCGTGACCCAGG---ACTCTCTCCCTGCAGGAGCGGTTCATCT 358
Db 158 -----ProProThrProLysProProValTyrThrProAsn----- 171
QY 359 ACAAGGTGAAGTTCATCGGCGTGAATTCCTCCGACG-----GCCCGGTGATGC 409
Db 172 -----ProLysProProValThrLysProProThrHisThrProSerPro----- 186
QY 410 AGAAGAGACCATGGTGGAGGCGCTCCACGAGCGCTGTACCCCGGCGGACGGCGTGC 469
Db 187 -----LysProProThrSerLysProProValTyrThrProSer----- 201
QY 470 TGAAGGCGGAGATCCACAGGCGCTGAAGTGAAGGAGCGCGGCGCACTACCTGGTGGAGT 529
Db 202 -----ProLysProProLysPro----- 207
QY 530 TCAAGTCCATCATCATG-----CCAAGAGCGCGTGCAGCTGCCCGGCTACTACT 580
Db 208 ---SerProProThrTyrThrProThrProLysProAlaThrLysProProThrSer 226
QY 581 ACGTGGACA----- 589
Db 227 ThrProThrHisProLysProThrProHisThrProTyrProGlnAlaHisProProThr 246
QY 590 -----CCAAGCTGGACATCACT---CCACACAGGAGACTACACCA 628
Db 247 TyrLysProAlaProLysProSerProProAlaProThrProProThrTyrThrProPro 266
QY 629 TCGTGGACCATGAGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 674
Db 267 ValSerHisThrProSerSerPro-ProProProProProProProProPro 281

RESULT 9
VGLX_HSVEB
ID VGLX_HSVEB STANDARD; PRT; 797 AA.
AC P28958;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN 71
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; M86664; AAB02506.1; -.
DR PIR; H36802; VGBEX1.
DR InterPro; IPR007110; Ig-like.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 797 GLYCOPROTEIN X.
FT DOMAIN 23 465 SER/THR-RICH.
FT TRANSMEM 766 790 POTENTIAL.
FT CARBOHYD 590 590 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 737 AA; 80342 MW; 50C9ED9211F5E5B2 CRC64;

Alignment Scores:
Pred. No.: 0.0226 Length: 797
Score: 151.50 Matches: 59
Percent Similarity: 46.50% Conservative: 34
Best Local Similarity: 29.50% Mismatches: 74
Query Match: 12.25% Indels: 33
DB: 1 Gaps: 6

US-10-081-864-14 (1-678) x VGLX_HSVB (1-797)
Qy 68 ACGCCACGAGTTCGAGATCGAGGCGAGGCGAGGCGCCCTCGAGGCGCAACA 127
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28 ThrThrThrSerSerSerThrSerGlySerGlyGlnSerThrSerSerGlyThr 47
Qy 128 CCGTGAAGTTGAAGTGACCAAGGCGCGCCCTCGCTCGGACATCTCTGCC 187
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
48 AsnSerSer-----SerSerProThrThrSerProPro 58
Qy 188 CCGAGTTCACGCTCCAGGTGACGTGAGGACACCGCGGACATCCCGACTACA 247
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
59 ThrThrSerSerProProThrSerThrHisThrSerSerProSerThrSerThr 78
Qy 248 AGAAGCTGTCTCTCCCGAGCGGCTTCAAGTGGGAGCGCGTGATCAACTCGAGACGGCG 307
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
79 GlnSer---SerSerThrAlaAlaThrSerSerAlaProSerThrAlaSerThr 97
Qy 308 CGTGGGCGCGTACCCAGGAGCTCTCTCGAGGAGCGCTCTCATCTACAAGTGA 367
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
98 Thr-----SerIleProThrSerThrSerThrGluThrThrThrThr 111
Qy 368 AGTTCATCGGCGTCACTTCCCTCCGACGCGCGCGTGATGACAGACCATGGCT 427
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
112 ThrProThrAlaSerThrThrProThrThr----- 123
Qy 428 GGGAGGCGCTCACGAGCGCGCTGTACCCCGGACGCGGTGTGAAGGCGGAGATCA 487
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
124 ---ThrAlaAlaProThrThrAlaAlaThrThrAlaValThrThrAlaAlaSerThr 142
Qy 488 AGGCGCTGAGCTGAAGGAGCGCGCGCCACATCTCGTGGAGTTCAGTCTCATCATGS 547
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
143 Ser-----AlaGluThrThrAlaThrAlaThrAlaThrSerThrProThrThr 159
Qy 548 CCAAGAGCGCGTGCAGCTGCCCGGCTACTACTGCGGACCAAGCTGGACATCACCT 607
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
160 -----ThrThrProThrSerThrThrThrThrAlaThrThrThrVal 174
Qy 608 CCACACGAGGAGTACCATCATCTGTGAGGAGTACGAGCGCGAGGCGCGCCGCCACC 667
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 ProThrThrAlaSerThrThrThrAspThrThrThrAlaAlaThrThrThrAlaAlaThr 194

RESULT 10
MUSB_HUMAN
ID MUSB_HUMAN STANDARD; PRT: 5703 AA.
AC Q9HC84; Q00447; Q00573; Q14985; Q15494; Q95291; Q95451; Q14881;
DC Q9552; Q9UE28;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 5B precursor (Mucin 5 subtype B, tracheobronchial) (High
DE molecular weight salivary mucin MGI) (Sublingual gland mucin).
GN MUSCB OR MUSC5.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-1594 FROM N.A.
RA Chen Y., Di Y.P., Wu R.;
RT "Molecular cloning of the amino-terminal and 5'-flanking region of the
RT human MUSCB mucin gene.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBAJ databases.
RN [2]
RP SEQUENCE OF 1-1325 FROM N.A.
RX MEDLINE=99009274; PubMed=9790959;
RA Offner G.D., Nunes D.P., Keates A.C., Afzal N.H., Troxler R.F.;
RT "The amino-terminal sequence of MUSCB contains conserved
RT multifunctional D domains: implications for tissue-specific mucin
RT functions.";
RL Biochem. Biophys. Res. Commun. 251:350-355(1998).
RN [3]
RP SEQUENCE OF 40-1324 FROM N.A.
RX MEDLINE=99029392; PubMed=9804771;
RA Desseyn J.-L., Buisine M.P., Porchet N., Aubert J.-P., Laine A.;
RT "Genomic organization of the human mucin gene MUSCB: cDNA and genomic
RT sequences upstream of the large central exon.";
RL J. Biol. Chem. 273:30157-30164(1998).
RN [4]
RP SEQUENCE OF 1326-4895 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97166151; PubMed=9013550;
RA Desseyn J.-L., Guyonnet-Duperat V., Porchet N., Aubert J.-P.,
RA Laine A.;
RT "Human mucin gene MUSCB, the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super-repeat. Structural
RT evidence for a lip15.5 gene family.";
RL J. Biol. Chem. 272:3168-3178(1997).
RN [5]
RP SEQUENCE OF 4057-4480 FROM N.A.
RC TISSUE=Salivary gland;
RX MEDLINE=97292540; PubMed=9147051;
RA Nielsen P.A., Bennett E.P., Wandall H.H., Therkildsen M.H.,
RA Hannibal J., Clausen H.;
RT "Identification of a major human high molecular weight salivary mucin
RT (MGI) as tracheobronchial mucin MUSCB.";
RL Glycobiology 7:413-419(1997).
RN [6]
RP SEQUENCE OF 4721-5703 FROM N.A.
RC TISSUE=Gall bladder;
RX MEDLINE=97293229; PubMed=9164870;
RA Keates A.C., Nunes D.P., Afzal N.H., Troxler R.F., Offner G.D.;
RT "Molecular cloning of a major human gall bladder mucin: complete C-
RT terminal sequence and genomic organization of MUSCB.";
RL Biochem. J. 324:295-303(1997).
RN [7]
RP SEQUENCE OF 4809-5687 FROM N.A.
RC TISSUE=Sublingual gland;
RX MEDLINE=96125355; PubMed=8554565;
RA Troxler R.F., Offner G.D., Zhang F., Iontcheva I., Oppenheim F.G.;
RT "Molecular cloning of a novel high molecular weight mucin (MGI)
RT from human sublingual gland.";
RL Biochem. Biophys. Res. Commun. 217:1112-1119(1995).
RN [8]
RP SEQUENCE OF 4859-5703 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97347489; PubMed=9201995;
RA Desseyn J.-L., Aubert J.-P., Porchet N., Laine A.;
RT "Genomic organization of the 3 region of the human MUSCB mucin.";
RL J. Biol. Chem. 272:16873-16883(1997).
CC -!- FUNCTION: Salivary mucin that is thought to contribute to the
CC lubricating and viscoelastic properties of whole saliva.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed mainly in bronchus glands and also
CC in submaxillary glands, endocervix, gall bladder, and pancreas.
CC -!- PTM: Highly glycosylated.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.

```

CC -!- SIMILARITY: Contains 3 WVF domains.
 CC -!- SIMILARITY: Contains 4 WVF domains.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF107890; AAC3673.1; -;
 CC EMBL; AF086604; AAC67545.1; -;
 CC EMBL; AJ004862; CAA06167.1; -;
 CC EMBL; Z72496; CAA96577.1; -;
 CC EMBL; X74955; CAA52910.1; -;
 CC EMBL; X74955; CAA52910.1; -;
 CC EMBL; U63836; AAB81398.1; -;
 CC EMBL; U78554; AAC51344.1; -;
 CC EMBL; U78552; AAC51344.1; JOINED.
 CC EMBL; U78553; AAC51344.1; JOINED.
 CC EMBL; U78551; AAC51343.1; -;
 CC EMBL; U95031; AAB65151.1; -;
 CC EMBL; Y09788; CAA70926.1; -;
 CC EMBL; Y09788; CAA70926.1; -;
 CC Genew; HGNC:7516; MUC5B.
 CC MIM; 600770; -;
 CC GO; GO:0005515; F:protein binding; IPI.
 CC InterPro; IPR006208; Cys knot.
 CC InterPro; IPR006207; Cys knot C.
 CC InterPro; IPR009041; BMP_inhibitor.
 CC InterPro; IPR002919; TIL_Cysrich.
 CC InterPro; IPR06552; VC_out.
 CC InterPro; IPR001007; WVF_C.
 CC InterPro; IPR001846; WVF_D.
 CC Pfam; PF00007; Cys knot; 1.
 CC Pfam; PF01826; TIL; 1.
 CC Pfam; PF00093; vwc; 1.
 CC Pfam; PF00094; vwd; 4.
 CC SMART; SM00214; VMC; 6.
 CC SMART; SM00215; VMC_out; 4.
 CC SMART; SM00216; VWD; 4.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01235; CTCK_2; 1.
 CC PROSITE; PS01208; WVF_1; 2.
 CC PROSITE; PS0184; WVF_2; 2.
 CC Glycoprotein; Repeat; Signal; Polymorphism.
 KW SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 5703 MUCIN 5B.
 FT DOMAIN 77 225 WVF 1.
 FT DOMAIN 329 386 TIL.
 FT DOMAIN 426 580 WVF 2.
 FT DOMAIN 858 918 WVF 1.
 FT DOMAIN 896 1044 WVF 3.
 FT DOMAIN 1457 1603 THR-RICH.
 FT DOMAIN 1609 4873 THR-RICH.
 FT DOMAIN 5005 5178 WVF 4.
 FT DOMAIN 5353 5425 WVF 2.
 FT DOMAIN 5462 5528 WVF 3.
 FT DOMAIN 5594 5683 CTCK.
 FT DISULFID 5594 5646 BY SIMILARITY.
 FT DISULFID 5622 5660 BY SIMILARITY.
 FT DISULFID 5626 5676 BY SIMILARITY.
 FT DISULFID 5643 5678 BY SIMILARITY.
 FT DISULFID 5645 5682 BY SIMILARITY.
 FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 516 516 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 806 806 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 930 930 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1277 1277 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1293 1293 N-LINKED (GLCNAC. . .)

FT CARBOHYD 1557 1557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1775 1775 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2192 2192 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2721 2721 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3419 3419 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3419 3419 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3948 3948 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4491 4491 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4901 4901 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4958 4958 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4965 4965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4987 4987 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5037 5037 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5052 5052 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5156 5156 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5427 5427 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5467 5467 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5506 5506 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5507 5507 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5543 5543 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5553 5553 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5604 5604 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5618 5618 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5662 5662 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 5137 5137 T -> S (in dbSNP:2672788).
 FT FTID=VAR_014123.
 FT G -> E (IN REF. 2).
 FT FPGCN -> LPCLCK (IN REF. 2).
 FT S -> C (IN REF. 2).
 FT E -> K (IN REF. 1).
 FT R -> S (IN REF. 2).
 FT PL -> T (IN REF. 2).
 FT E -> N (IN REF. 2).
 FT E -> K (IN REF. 2).
 FT E -> R (IN REF. 2).
 FT MISSING (IN REF. 2 AND 3).
 FT D -> N (IN REF. 2).
 FT RT -> TR (IN REF. 2).
 FT RK -> GR (IN REF. 2).
 FT L -> P (IN REF. 2).
 FT GAA -> AH (IN REF. 3).
 FT A -> S (IN REF. 3).
 FT DP -> RS (IN REF. 2).
 FT F -> L (IN REF. 2).
 FT A -> P (IN REF. 3).
 FT 676

Alignment Scores:

Pred. No.: 0.0286 Length: 5703
 Score: 150.50 Matches: 73
 Percent Similarity: 39.67% Conservative: 23
 Best Local Similarity: 30.17% Mismatches: 85
 Query Match: 12.17% Indels: 61
 DB: 1 Gaps: 13

US-10-081-864-14 (1-678) x MUSB_HUMAN (1-5703)

QY 68 ACGGCCAGAGTTCGAGATCGAGGCGGAGGCGGCGCCCT-----ACGAGG 118
 Db 2060 ThrAlaThrProSerSerProGlyArgAlaargThrLeuProValTrpIleSerThr 2079
 QY 119 GCCCAACACCGTGAAGTTGAAGTGACCAAGGCGGCGCCCTCCCTCCCTGGGACA 178
 Db 2080 ThrThrThrPro---ThrThrArgGlySerThrValThrProSerSerIleProGlyThr 2098
 QY 179 TCC-----TGT 184
 Db 2099 ThrHisThrProThrValLeuThrThrThrThrValAlaThrGlySerMetAla 2118
 QY 185 CCCCCCAGTTCAGTACGGTCCCAAGGTGAGTGAAGCAGCCGCGCCGACATCCCCGACT 244
 Db 2119 ThrProSerSerSerThrGln-----ThrSerGlyThrProProSerLeuThrThr 2135
 QY 245 ACAAGAAGCTGCTCTTCCCGAGGCGTCAAGTGGAGCGGTGATGAACATTCGAGGACG 304

D5 2136 ThrAlaThrThrIleThr-----AlaThrGlySerThrThrAsnProSerSerThr 2152
 QY 305 GCGGCGTGGGACCGTGCAGCCAGGACCTCCCTCGCAGGACGGTCTTCATCTACAAGG 364
 D5 2153 ProGlyThrThrPro-----IleProProValLeuThrThrAlaThrThrPro 2169
 QY 365 TGAAGTTCATCGCGTGAACCTCCCTCGCAGGACGGTCTTCATCTACAAGG 424
 D5 2170 AlaAlaThrSerThrValThrProSerSerAlaLeuGlyThrThrHisThrPro--- 2188
 QY 425 GCTGGGAGGCTCCACCGAGCGCTGTACCCCGCAGGCGGCTGCTCAAGGGGAGATCC 484
 D5 2189 -----ProValProAsn-----ThrThrAlaThrThrHisGlyA-GSer 2201
 QY 485 ACAAGGCGCTGAAGCTGAAGAGCGGCGGCGCACTACCTGGTGGAGTTCAAGT----- 535
 D5 2202 LeuSerProSerProHisThrValCysThrAlaThrThrSerAlaThrSerGlyIle 2221
 QY 536 -----COATCTACATGTCGCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571
 D5 2222 LeuGlyThrThrHisIleThrGluProSerThr-----GlyThrSerHisThrPro 2238
 QY 572 GCTACTACTACTCTGG-----ACCAAGCTGGGACATCACCT---CCACA 613
 D5 2239 AlaAlaThrThrGlyThrThrGlnHisSerThrProAlaLeuSerProHisProSer 2258
 QY 614 ACGAGGACTACA-----CCATCGTGGAGGAGTACGAGCGGCGGCGGCGGCGGCGG 667
 D5 2259 SerArgThrThrGluSerProProSerProGlyThrThrThrProGlyHisThrAla 2278
 QY 668 TGTTCC 673
 D5 2279 ThrSer 2280
 RESULT 11
 SFPO_HUMAN
 ID_SFPO_HUMAN STANDARD; PRT; 707 AA.
 AC P23246; P30808; PRT;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-
 binding protein-associated splicing factor) (PTB-associated splicing
 factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
 DE SFPO OR PSF
 GN Homo sapiens (Human)
 OS
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
 RC TISSUE=Fetal brain;
 RX MEDLINE=93194059; PubMed=8443401;
 RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
 RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
 factor.";
 RL Genes Dev. 7:393-406(1993).
 RN [2]
 RP SEQUENCE OF 312-707 FROM N.A.
 RC TISSUE=Fetal skeletal muscle;
 RX MEDLINE=90091812; PubMed=2480877;
 RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh P.S.;
 RT "Cloning and characterization of a myoblast cell surface antigen
 defined by 24.1D5 monoclonal antibody.";
 RL Development 105:723-731(1998).
 RN [3]
 RP SEQUENCE OF 48-68 AND 213-246.
 RX MEDLINE=93176127; PubMed=8439294;
 RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
 RT "Purification and characterization of a DNA-binding heterodimer of 52
 and 100 kDa from HeLa cells.";
 RL Biochem. J. 290:267-272(1993).
 CC -!- FUNCTION: Essential pre-mRNA splicing factor required early in

spliceosome formation. Binds to the mammalian polypyrimidine
 tracts. Forms a complex with the polypyrimidine tract-binding
 protein (PTB). Seems to also bind DNA.
 CC -!- SUBUNIT: Heterotrimer of two 52 kDa and two 100 kDa subunits.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=Long;
 CC IsoId=P23246-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P23246-2; Sequence=VSP_005855;
 CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
 CC -!- CAUTION: Was originally (Ref.2) thought to be myoblast cell
 CC surface antigen 24.1D5 and a possible membrane-bound protein
 CC ectokinase.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X70944; CAA50283.1; -;
 CC EMBL; X16850; CAA34747.1; -;
 CC PIR; A46302; A46302.
 CC PIR; S29770; S29770.
 CC HSSP; P11940; 1CVJ.
 CC SWISS-2DPAGE; P23246; HUMAN.
 CC Genew; HGNC:10774; SFPQ.
 CC GK; P23246; -;
 CC MIM; 605199; -;
 CC GO; GO:0008248; P:pre-mRNA splicing factor activity; TAS.
 CC GO; GO:0006397; P:mRNA processing; TAS.
 CC GO; GO:0006371; P:mRNA splicing; TAS.
 CC InterPro; IPR000504; RNA_rec_mot.
 CC Pfam; PF00076; Rnm; 2.
 CC SMART; SM00360; RRM; 2.
 CC PROSITE; PS50102; RRM; 2.
 CC PROSITE; PS00030; RRM_RNP_1; 1.
 CC Nuclear protein; RNA-Binding; DNA-binding; mRNA splicing; Repeat;
 KW Alternative splicing.
 KW DOMAIN 297 369 RNA-BINDING (RRM) 1.
 FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
 FT DOMAIN 9 27 3 X 3 AA REPEATS OF R-G-G.
 FT REPEAT 9 11 1.
 FT REPEAT 19 21 2.
 FT REPEAT 25 27 3.
 FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
 FT DOMAIN 10 15 POLY-GLY.
 FT DOMAIN 20 27 POLY-GLY.
 FT DOMAIN 56 65 POLY-PRO.
 FT DOMAIN 67 71 POLY-GLN.
 FT DOMAIN 95 98 POLY-GLN.
 FT DOMAIN 99 103 POLY-PRO.
 FT DOMAIN 184 188 POLY-ARG.
 FT DOMAIN 571 574 POLY-ARG.
 FT DOMAIN 613 616 POLY-GLY.
 FT DOMAIN 635 641 POLY-GLY.
 FT VARSPLIC 663 707
 FT
 FT KPRF -> VRMIDVG (in isoform Short).
 FT /FTID=VSP_005855.
 FT G -> R (IN REF. 3).
 FT CONFLICT 243 243 6D8D5EA95E235847 CRC64;
 SQ SEQUENCE 707 AA; 76149 MW; 6D8D5EA95E235847 CRC64;
 Alignment Scores:
 Pred. No.: 0.0278
 Score: 150.00
 Percent Similarity: 35.09%
 Best Local Similarity: 29.39%
 Query Match: 12.13%
 Mismatches: 65
 Indels: 84
 Matches: 67
 Length: 707

DB:	1	Gaps:	14
US-10-081-864-14 (1-678) x SFPQ_HUMAN (1-707)			
QY	105	CGGCGCTACGAGGCGCAACACACCGTGAAGTTGAAGTGAACAGGCGGCGCCCTGC-	163
Db	83	ProProHisGlnProProHisProGlnProHisGlnGlnGlnProProProPro 102	
QY	164	-----CCTTCGCTGGGACA---	178
Db	103	ProGlnAspSerSerLysProValAlaGlnGlyProGlyProAlaProGlyValGly 122	
QY	179	TCTGTCTCCCGAGTCCAGTACGCTCAAGGTGTACGTGAAGCACCCTCCGACATCC 238	
Db	123	SerAlaProProAlaSerSerAla-----ProProAlaThr 135	
QY	239	CGGACTACAGAAGCTGCTCTCCCGAGGCTTCAAGTGGAGCGCGTGAATCTCG 298	
Db	136	ProProThrSerGlyAlaPro--ProGlySerGlyProGly----- 148	
QY	299	AGGAGCGGCGTGGCGGCGGACCGTGAACCCAGGACTCTCTCCCTGAGGAGCGTCTCT 352	
Db	149	-----ProThrProThrProProProAlaValThrSerAlaProPro 162	
QY	353	-----TCATCTACAAGGTGAAGTTCATCGGCGTGAATCTCCCTCCGACGCGCCCG 403	
Db	163	GlyAlaProProProProProSerSerGlyValProThrThrProProGlnAla-- 181	
QY	404	TGATCAGAGAAGACCATGGCGTGGAGCGCTCCACGAGCGCTGTACCCCGCGAGC 463	
Db	182	-----GlyGlyProProProProProAlaValProGlyPro 194	
QY	464	CGCTGCT-----GAAGGCGGATCCACAGGCGCTGAAGCTGAAGGCGCGGCG 514	
Db	195	Gly-ProGlyProGlyGlnGlyProGlyProGlyProGlyGly--GlyLysMetPr 213	
QY	515	A-----CTACTGTGGTGA-----GTTCA 532	
Db	213	OGlyGlyProLysProGlyGlyProGlyLeuSerThrProGlyGlyHisProLysPr 233	
QY	533	AGTCCATCTACATGGCGCAAGACCGCTGCAGCTGCCGGCTACTACTACTGACCA 592	
Db	233	oProHisArgGlyGlyGlyProArgly-----GlyArgG 246	
QY	593	AGCTGACATCACTCCCAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCG 652	
Db	246	n-----HisHisProProTyHisGlnGlnHis-----HisG 257	
QY	653	AGGCGCGCCACCATCTTCTCT 674	
Db	257	nGlyProProProGlyGlyPro 264	
RESULT 12			
IF2_STRAW		STANDARD;	PRT; 1046 AA.
ID	IF2_STRAW		
AC	Q82K53;		
DT	15-MAR-2004 (Rel. 43, Created)		
DT	15-MAR-2004 (Rel. 43, Last sequence update)		
DE	Translation initiation factor IF-2.		
GN	INF2 OR SAV2551.		
OS	Streptomyces avermitilis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxID=33903;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;		
RX	MEDLINE=21477403; PubMed=11572948;		
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,		
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,		
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.,		
RT	"Genome sequence of an industrial microorganism Streptomyces		

avermittilis: deducing the ability of producing secondary metabolites.";			
Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).			
(2)			
SEQUENCE FROM N.A.			
STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;			
MEDLINE=22608306; PubMed=12692562;			
Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,			
Sakaki Y., Hattori M., Omura S.;			
"Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis.";			
Nat. Biotechnol. 21:526-531(2003).			
-!- FUNCTION: One of the essential components for the initiation of protein synthesis. Protects formylmethionyl-tRNA from spontaneous hydrolysis and promotes its binding to the 30S ribosomal subunits. Also involved in the hydrolysis of GTP during the formation of the 70S ribosomal complex (By similarity).			
-!- SUBCELLULAR LOCATION: Cytoplasmic.			
-!- SIMILARITY: Belongs to the IF-2 family.			
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
EMBL; AP005031; BAC70262.1; --			
HAWAP; MF_00100; -- 1.			
InterPro; IPR001851; Bac_inmem_transp.			
InterPro; IPR004161; EFTU_D2.			
InterPro; IPR000795; EF_GTPbind.			
InterPro; IPR003577; GTPase_Ras.			
InterPro; IPR000178; IF2.			
InterPro; IPR006847; IF2_N.			
InterPro; IPR002965; P-rich_extensn.			
InterPro; IPR001806; Ras_trnsfrmg.			
InterPro; IPR005225; Small_Grp.			
InterPro; IPR009000; Translat_factor.			
Pfam; PF00009; GTP_EFTU; 1.			
Pfam; PF03144; GTP_EFTU_D2; 1.			
Pfam; PF04760; IF2_N; 2.			
PRINTS; PR00315; ELONGATNFCT.			
PRINTS; PR01217; PRICHEXTENS.			
PRINTS; PR00449; RASTRNSFRMG.			
ProDom; PD186100; IF2; 1.			
SMART; SM00173; Ras; 1.			
TIGRFAMs; TIGR00487; IF-2; 1.			
TIGRFAMs; TIGR00231; small_GTP; 1.			
PROSITE; PS01176; IF2; FALSE_NEG.			
KW Initiation factor; Protein biosynthesis; GTP-binding;			
KW Complete proteome.			
FT DOMAIN 542 694			
FT NP_BIND 548 555			
FT NP_BIND 598 602			
FT NP_BIND 652 655			
SQ SEQUENCE 1046 AA; 106882 MW; 62B12E2DFA3AA595 CRC64;			
Alignment Scores:			
Pred. No.:	0.0284	Length:	1046
Score:	150.00	Matches:	69
Percent Similarity:	38.64%	Conservative:	16
Best Local Similarity:	31.36%	Mismatches:	91
Query Match:	12.00%	Indels:	44
DB:	1	Gaps:	12
US-10-081-864-14 (1-678) x IF2_STRAW (1-1046)			
QY	655	CCTCGGTGCGCTCGTACTGCTCCAGCTGGTAGTCTCGTGTGGGGGATGTCCA 596	
Db	101	ProAlaAlaAlaProAlaAlaPro-----GlyProArgProGly-----Pro 114	

Db 1392 SerProThrThrSerThrProSerSerThrProGlnProThr-----SerSer 1408
QY 227 CCGCGGACATCCCGGACTACAGAAGCTGTCTTCCCGAGGGCTCAAGTGGGAGCGG 286
Db 1409 ProThrThrLeuProThrThr-----SerProLeuThrSerSerAlaThrSer 1424
QY 287 TGATGAACTCCGAGGACGCGCGCTGGCGACCGTGCAGGACCTCCCTGCGAGGACG 346
Db 1425 -----ProThrThrSerHisIleThrSer 1432
QY 347 GCTGCTTCATCTACAGGTGAAGTTTCATCGGGTGAAGTCTCCCGAGCGCCCGGTGA 406
Db 1433 ThrValSerProThrThrSerProThrThrSerProThrThrSerPro--- 1451
QY 407 TCGAGAGAAGACCATGGCTGGAGGCTCCACCGAGCGCTGTACCCCGCGAGCGG 466
Db 1452 -----ThrThrSerThr--- 1455
QY 467 TCGTGAAGGCGGAGATCCACAGGCGCTGAAGCTGAAGGACGCGCGGACCTACCTCGGTGG 526
Db 1456 -----ThrSerPro-----ThrThrSerThrThrSerPro 1465
QY 527 AGTTCAGTCCATCTACATGCGGACGAGAGCCGTCGAGTCCCGGCTACTACTAGGTGG 586
Db 1466 ThrProSerProThrThrSerThrThrSerProThrProSerProThrThrSerThr 1485
QY 587 ACACCAAGCTGGACATCCACTCCACACAGGAGCTACACCATCGTGGAGCAGTACGAGC 646
Db 1486 SerProThr---ProSerProThrThrSerThrThrSerProThrThrSerProIleThr 1504
QY 647 GCACCGAGGCGCCACCCACC 667
Db 1505 SerProThrThrSerThrThr 1511

RESULT 14
SFRG_HUMAN
ID SFRG_HUMAN STANDARD; PRT; 659 AA.
AC Q8N2M8; O96026; Q96DX2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Splicing factor, arginine/serine-rich 16 (Suppressor of white-apricot
DE homolog 2).
GN SFRS16 OR SWAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP YOSHIMURA K., MURRAY J.C.;
RA "A transcriptional map in the region of 19q13 derived using direct
RT sequencing and exon trapping";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA YOSHIMURA K., MURRAY J.C.;
RT "NEDO human cDNA sequencing project";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein V.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
Rana S.S., Iqbal N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Besak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.B.,
Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Probably functions as an alternative splicing regulator.
CC May regulate the mRNA splicing of genes such as CLK1. May act by
CC regulating members of the CLK kinase family (by similarity).
CC -!- SUBUNIT: Probably interacts with CLK4 (by similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8N2M8-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8N2M8-2; Sequence=VSP_008207, VSP_008208, VSP_008209,
CC VSP_008210;
CC Note=No experimental confirmation available;
CC -!- PTM: Phosphorylated in vitro by CLK4 (by similarity).
CC -!- SIMILARITY: Belongs to the splicing factor SR family.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to some
CC missing sequence in an exon that changes the coding frame.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; AF042800; AAC82339.1; --
EMBL; AF042810; AAC82340.1; --
EMBL; AF042802; AAC82340.1; JOINED.
EMBL; AF042803; AAC82340.1; JOINED.
EMBL; AF042804; AAC82340.1; JOINED.
EMBL; AF042805; AAC82340.1; JOINED.
EMBL; AF042806; AAC82340.1; JOINED.
EMBL; AF042807; AAC82340.1; JOINED.
EMBL; AF042808; AAC82340.1; JOINED.
EMBL; AF042809; AAC82340.1; JOINED.
EMBL; AK074590; BAC11078.1; --
EMBL; BC013178; AAH13178.1; ALT_SEQ.
Genew; HGNC:17731; SFRS16.
KW mRNA processing; mRNA splicing; Nuclear protein; Alternative splicing;
KW Polymorphism.
FT DOMAIN 354 658 ARG-RICH.
FT DOMAIN 362 520 SER-RICH.
FT VARSPLIC 1 334 Missing (in isoform 2).
FT VARSPLIC 335 359 /FTId=VSP_008207.
FT VARSPLIC 456 493 GVTGKPPAPQPGGPPAPGRNASAR -> MGTGRRGGPRL
FT VARSPLIC 493 SRSHGDEYRGGRLRHSSSRSSWSLSPSRSL ->
FT CWNLCVEVGAGGARSPIGRCSEGGNLAQLREHPI
FT (in isoform 2).
FT /FTId=VSP_008209.
FT VARSPLIC 494 659 Missing (in isoform 2).
FT VARSPLIC 159 159 E -> K (in dbSNP:4803794).
FT VARIANT 410 410 S -> P (IN REF. 2).
FT CONFLICT 659 AA; 75214 MW; 1E455FA232BF30D4 CRC64;
SQ SEQUENCE

344 ACGGCTGCTTCATCTACAAGGTGAAGTTCATCGCGTGAAGTTCCTCCGACGGCCCG 403
80 SerThrAlaSerThrThr-----SerGlyProGlyThrThrPro 92
404 TGATGACAGAAGAACATCGGCTGGAGGCTCCACCGAGCGCTGTACCCCGGAGC 463
93 -----SerProValProThrThrSerThrThrSer--- 102
464 GCGTGTGAAGGCGGATCCACAGGCCCTGAAGCTGAAGGCGGCGGACCTACCTGG 523
103 -----AlaProThrThrSerThrThrSerAlaSerThrAlaSerThrThr--- 117
524 TGGAGTTCAGAGTCCATCTACATCGCCCAAGAGCCCGTGCAGCTGCCGGTACTACTACG 583
118 -----SerGlyProGlyThr-----SerLeuSerProValProThrThrSerThr 132
584 TGGACACCAAGCTGGAGATCACCTCCACACAGGAGGACTACA----- 625
133 -----ThrSerAlaProThrThrSerThrThrSerGlyProGlyThrThr 147
626 -----CCATCGTGGAGCAGTACGAGCCACCGAGGCGGCGGACCCACC 667
148 ProSerProValProThrThrSerThrThrSerAlaProThrThrSerThrThr 165

Search completed: July 29, 2004, 14:33:24
Job time : 20.7906 secs

693 697 702 1233
698 707 69
703 707 21
708 712 56
709 712 12
710 712 12
711 712 12
712 712 12
713 712 12
714 712 12
715 712 12
716 712 12
717 712 12
718 712 12
719 712 12
720 712 12
721 712 12
722 712 12
723 712 12
724 712 12
725 712 12
726 712 12
727 712 12
728 712 12
729 712 12
730 712 12
731 712 12
732 712 12
733 712 12
734 712 12
735 712 12
736 712 12
737 712 12
738 712 12
739 712 12
740 712 12
741 712 12
742 712 12
743 712 12
744 712 12
745 712 12
746 712 12
747 712 12
748 712 12
749 712 12
750 712 12
751 712 12
752 712 12
753 712 12
754 712 12
755 712 12
756 712 12
757 712 12
758 712 12
759 712 12
760 712 12
761 712 12
762 712 12
763 712 12
764 712 12
765 712 12
766 712 12
767 712 12
768 712 12
769 712 12
770 712 12
771 712 12
772 712 12
773 712 12
774 712 12
775 712 12
776 712 12
777 712 12
778 712 12
779 712 12
780 712 12
781 712 12
782 712 12
783 712 12
784 712 12
785 712 12
786 712 12
787 712 12
788 712 12
789 712 12
790 712 12
791 712 12
792 712 12
793 712 12
794 712 12
795 712 12
796 712 12
797 712 12
798 712 12
799 712 12
800 712 12
801 712 12
802 712 12
803 712 12
804 712 12
805 712 12
806 712 12
807 712 12
808 712 12
809 712 12
810 712 12
811 712 12
812 712 12
813 712 12
814 712 12
815 712 12
816 712 12
817 712 12
818 712 12
819 712 12
820 712 12
821 712 12
822 712 12
823 712 12
824 712 12
825 712 12
826 712 12
827 712 12
828 712 12
829 712 12
830 712 12
831 712 12
832 712 12
833 712 12
834 712 12
835 712 12
836 712 12
837 712 12
838 712 12
839 712 12
840 712 12
841 712 12
842 712 12
843 712 12
844 712 12
845 712 12
846 712 12
847 712 12
848 712 12
849 712 12
850 712 12
851 712 12
852 712 12
853 712 12
854 712 12
855 712 12
856 712 12
857 712 12
858 712 12
859 712 12
860 712 12
861 712 12
862 712 12
863 712 12
864 712 12
865 712 12
866 712 12
867 712 12
868 712 12
869 712 12
870 712 12
871 712 12
872 712 12
873 712 12
874 712 12
875 712 12
876 712 12
877 712 12
878 712 12
879 712 12
880 712 12
881 712 12
882 712 12
883 712 12
884 712 12
885 712 12
886 712 12
887 712 12
888 712 12
889 712 12
890 712 12
891 712 12
892 712 12
893 712 12
894 712 12
895 712 12
896 712 12
897 712 12
898 712 12
899 712 12
900 712 12
901 712 12
902 712 12
903 712 12
904 712 12
905 712 12
906 712 12
907 712 12
908 712 12
909 712 12
910 712 12
911 712 12
912 712 12
913 712 12
914 712 12
915 712 12
916 712 12
917 712 12
918 712 12
919 712 12
920 712 12
921 712 12
922 712 12
923 712 12
924 712 12
925 712 12
926 712 12
927 712 12
928 712 12
929 712 12
930 712 12
931 712 12
932 712 12
933 712 12
934 712 12
935 712 12
936 712 12
937 712 12
938 712 12
939 712 12
940 712 12
941 712 12
942 712 12
943 712 12
944 712 12
945 712 12
946 712 12
947 712 12
948 712 12
949 712 12
950 712 12
951 712 12
952 712 12
953 712 12
954 712 12
955 712 12
956 712 12
957 712 12
958 712 12
959 712 12
960 712 12
961 712 12
962 712 12
963 712 12
964 712 12
965 712 12
966 712 12
967 712 12
968 712 12
969 712 12
970 712 12
971 712 12
972 712 12
973 712 12
974 712 12
975 712 12
976 712 12
977 712 12
978 712 12
979 712 12
980 712 12
981 712 12
982 712 12
983 712 12
984 712 12
985 712 12
986 712 12
987 712 12
988 712 12
989 712 12
990 712 12
991 712 12
992 712 12
993 712 12
994 712 12
995 712 12
996 712 12
997 712 12
998 712 12
999 712 12
1000 712 12

Alignment Scores:
Pred. No.: 0.0507 Length: 1233
Score: 145.00 Matches: 69
Percent Similarity: 41.28% Conservative: 21
Best Local Similarity: 31.65% Mismatches: 56
Query Match: 11.80% Indels: 72
DB: 1 Gaps: 12

US-10-081-864-14 (1-678) x MUS_HUMAN (1-1233)

68 ACGGCGAGGTCGAGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 127
2 ThrSerThrSerAlaSerThrThrSerThrIleSerProLeuThrThrSerThrThr 21
128 CCGTGAAGTTGAAGTCAACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 175
22 -----SerAlaProIleThrSerMetProSerGlyProGly 33
176 ACATCTGTCCCGCCAGTTCAGTACGCTCCAGGTCGAGGTCGAGGTCGAGGTCGAGG 235
34 ThrThrProSerProValProThrThrSerProValProThrThrSerThrThr 50
236 TCC-----CCGACTACAGAGGCTGCTCCCGGAGGCGGCTTCAAGTGGGAGC 283
51 SerThrThrSerGlyProGlyThrThrProSerProValProThrThrSerThr 70
284 GCGTATCAACTTCAGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 343
71 AlaProThrThrSerThrSerAla----- 79

Blank sheet USPTO

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:27:25 ; Search time 40.3863 Seconds
(without alignments)
10593.763 Million cell updates/sec

Title: US-10-081-864-14
Perfect score: 1237
Sequence: 1 atggcctctccgagaagt.....gacaccactgttctgttaa 678

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DRV=xlp
-Q=cgna2_1/USPTO.spool.p/US10081864/runat_29072004_150506_25475/app.query.fasta_1.1742
-DB=SPTREMBL 25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864 @CGN_1_1_122 @runat_29072004_150506_25475 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	1191	96.3	225 5 Q9U6Y8	Q9U6Y8 discosoma s

Db 181 TyMetValLysProSerValGlnLeuProGlyTyrTyrValAspSerLysLeu 200
Qy 598 GACATCACTCCACACAGGAGGACTACCATCGTGAGCAGTACGAGCGACCGAGGC 657
Db 201 AspMetThrSerHisAsnGluAspTyrThrValValGluGlnTyrGluLysThrGlnGly 220
Qy 658 CGCACCACTGTCCTG 675
Db 221 ArgHisHisProPheille 226

RESULT 3

Q95P04 PRELIMINARY; PRT; 221 AA.
AC Q95P04; 221 AA.
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Gonopora tenuidens.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Fungina; Poritidae; Gonopora.
OX NCBI_TaxID=75301;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21538626; PubMed=11682051;
RA Gurskaya N.G., Fradkov A.F., Tersikh A., Matz M.V., Labas Y.A.,
Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
RT proteins(1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF381156; AAL27542.1; -
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 221 AA; 24918 MW; 93F9F4B5C2003CB4 CRC64;

Alignment Scores:

Pred. No.: 7,37e-51 Length: 221
Score: 795.00 Matches: 144
Percent Similarity: 80.48% Conservative: 25
Best Local Similarity: 68.57% Mismatches: 41
Query Match: 64.27% Indels: 0
DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x Q95P04 (1-221)

Qy 16 AACGTATCAACAGGCGCCCTTCAGTCCGATGAGGCGACCGTGAACGCCAC 75
Db 2 SerValileAlaLysGlnMetThrTyrLysValTyrMetSerGlyThrValAsnGlyHis 21
Qy 76 GAGTTCGAGATCGAGGCGAGGCGCGCCCTACGAGGCGCACACACCGTGAAG 135
Db 22 TyrPheGluValGlnGlyAspGlyLysGlyLysProTyrGluGluGlnThrValLys 41
Qy 136 TTGAAGGTGACCAAGGCGCCCTTCGCTTTCGCTGGGACATCTGTCGCCCGAGTTC 195
Db 42 LeuThrValThrLysGlyGlyProLeuProPheAlaTyrAspIleLeuSerProGlnSer 61
Qy 196 CAGTACGCTCCAAAGGTGACGTGACGACCCCGCGCATCCCGACTACAGAGCTG 255
Db 62 GlnTyrGlySerIleProPheThrLysTyrProGluAspIleProAspTyrValLysGln 81
Qy 256 TCCTTCCCGAGGCGCTTCAAGTGGGAGCGGTGATGAACCTTCGAGGACGCGCGTGGCG 315
Db 82 SerPheProGluGlyTyrThrTyrGluArgIleMetAsnPheGluAspGlyAlaValCys 101
Qy 316 ACCTGACCCAGACTCTCCCTCCGAGGCGGTCTTCACTCAAGGTGAAGTTCATC 375
Db 102 ThrValSerAsnAspSerIleGlnGlyAsnCysPheIleTyrAsnValLysPheSer 121

Qy 376 GCGGTGAACCTTCCTCCGAGCGCCCGTGTATGAGAGACCATGGCTGGAGGCC 435
Db 122 GlyLeuAsnPheProProAsnGlyProValMetGlnLysLysThrGlnGlyTyrGluPro 141
Qy 436 TCCACCGAGCGCTGTACCCCGCGAGCGCTGTGAAGCGCGAGATCCACAGGCCCTG 495
Db 142 AsnThrGluArgLeuPheAlaArgAspGlyMetLeuIleGlyAsnAsnPheMetAlaLeu 161
Qy 496 AAGCTGAAGGACGCGCGCCACTACTGCTGTGAGTTCAAGTCCATCTACATGCCAAGAG 555
Db 162 LysLeuGluGlyGlyGlyHisTyrLeuCysGluPheLysSerThrTyrLysAlaLysLys 181
Qy 556 CCGGTGACGTCCCGCGCTACTACTGCTGACACCAAGCTGGACATCACCTCCCAAC 615
Db 182 ProValLysMetProGlyTyrHisTyrValAspArgLysLeuAspValThrAsnHisAsn 201
Qy 616 GAGGACTACACATCGTGAGCAGTACGAG 645
Db 202 IleAspTyrThrSerValGluGlnCysGlu 211

RESULT 4

Q9U6Y7 PRELIMINARY; PRT; 232 AA.

ID Q9U6Y7
AC Q9U6Y7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fluorescent protein FP483.
OS Discosoma striata.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=105400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436614; PubMed=10504696;
RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
Markelov M.L., Lukyanov S.A.;
RT "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973(1999).
DR EMBL; AF168420; AAF03370.1; -
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFPLORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 232 AA; 26435 MW; AA8F18EE283CE4D CRC64;

Alignment Scores:

Pred. No.: 9.15e-46 Length: 232
Score: 726.50 Matches: 132
Percent Similarity: 76.00% Conservative: 39
Best Local Similarity: 58.67% Mismatches: 53
Query Match: 58.73% Indels: 1
DB: 5 Gaps: 1

US-10-081-864-14 (1-678) x Q9U6Y7 (1-232)

Qy 1 ATGGCTCTCCGAGACGTTCATCCAGTTCATCGCTTCAAGGTGCGCATGAGGCG 60
Db 1 MetSerCysSerLysSerValileLysGluGluMetLeuLeuAspLeuHisLeuGluGly 20
Qy 61 ACCGTGAACGCGCCAGCTTCGAGATCGAGCGGAGGCGGCGCCCTCGCTGGGACATC 120
Db 21 ThrPheAsnGlyHisTyrPheGluIleLysGlyLysGlyGlnProAsnGluGly 40
Qy 121 CACACACCGTGAAGTTCAGGTGACCAAGGCGCGCCCTCGCTTTCGCTGGGACATC 180
Db 41 ThrAsnThrValThrLeuGluValThrLysGlyGlyProLeuProPheGlyTyrHisIle 60
Qy 181 CTGTCCCCCGAGTTCAGTACGCTCCAGGTGACGTGAAGACCCCGCGCATATCCCC 240
Db 61 LeuCysProGlnPheGlnTyrGlyAsnLysAlaPheValHisHisProAspAsnIleHis 80

```

QY 241 GACTACAAAGAGCTGTCCTTCCCGAGCGGCTTCAAGTGGGAGCGCGTGAATCACTTCGAG 300
Db 81 AspTyrLeuLysLeuSerPheProGluGlyTyrThrTrpGluArgSerMetHisPheGlu 100
QY 301 GACGGCGCGTGGCGAGCGGTGACCCAGCACTCTCCCTCCGACCGCCCGGTGATGCAAGAAC 360
Db 101 AspGlyGlyLeuCysCysIleThrAsnAspIleSerLeuThrGlyAsnCysPheTyrTyr 120
QY 361 AAGTGAAGTTCATCGCGTGAATTCCTCCCTCCGACCGCCCGGTGATGCAAGAAC 420
Db 121 AspileLysPheThrGlyLeuAsnPheProAsnGlyProValValGlnLysLysThr 140
QY 421 ATGGCGTGGAGGCTCCACCGAGCGCTGTACCCCGCGACGCGTGTGAAGGCGGAG 480
Db 141 ThrGlyTrpGluProSerThrGluArgLeuTyrProArgAspGlyValLeuIleGlyAsp 160
QY 481 ATCCACAAAGCGCTGAAGTGAAGACGCGCGGCACTACTCTGTGGAGTTCAGAGTCCATC 540
Db 161 IleHisAlaLeuThrValGluGlyGlyHisTyrAlaCysAspIleLysThrVal 180
QY 541 TACATGCCCAAGAG---CCCGTCACTGCCCGGCTACTACTACGTGGACACCAAGCTG 597
Db 181 TyrArgAlaLysLysAlaLeuLysMetProGlyTyrHisTyrValAspThrLysLeu 200
QY 598 GACATCACTCCCAACAGGAGTACACCATCGTGGAGCAGTACGACGCGACCGAGGGC 657
Db 201 ValIleTrpAsnAspLysGluPheMetLysValGluGluHisGluIleAlaValAla 220
QY 658 CCGCACCACTGTTTC 672
Db 221 ArgHisLysProPhe 225

RESULT 5
Q963F5 PRELIMINARY; PRT; 225 AA.
ID Q963F5
AC Q963F5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Green fluorescent protein.
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Lesser M.P., Barry T.M., Mazel C., Matz M.V., Lukyanov S.A.,
RA Falkowski P., Gorbunov M., Kolber Z.;
RT "Green fluorescent proteins in Caribbean Scleractinian corals.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF384683; AAK62982.2; -
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 225 AA; 25847 MW; 77DE7D7C616929AF CRC64;

Alignment Scores:
Pred. No.: 1,31e-42 Length: 225
Score: 684.00 Matches: 119
Percent Similarity: 74.77% Conservative: 44
Best Local Similarity: 54.59% Mismatches: 55
Query Match: 55.30% Indels: 0
DB: 5 Gaps: 0

```

US-10-081-864-14 (1-678) x Q963F5 (1-225)

```

QY 76 GAGTTCGAGATCGAGGCGGAGCGCGCCCTACGAGGCGCCACAAACCGCTGAAG 135
Db 22 LysPheValIleLysGlyGluGlyGluLysProGluGlyThrGlnThrIleAsn 41
QY 136 TTGAAGGTGACAGAGGCGCGCCCTCCCTTCCTCGCTGGGACATCTCTGTCCTCC 195
Db 42 LeuThrValLysGluGlyAlaProLeuProPheAlaTyrAspIleLeuThrSerAlaPhe 61
QY 196 CAGTACGGCTCAAGTGTACGTGAGCACCACCGCCGACATCCCGCTACTCAAGAC 255
Db 62 GlnTyrGlyAsnArgValPheThrLysTyrProAspIleProAspTyrPheLysGln 81
QY 256 TCCTTCCCGAGGCGTTCAGTGGGAGCGCGTGTAGTAACTTCGAGGACGCGCGTGG 315
Db 82 ThrPheProGluGlyTyrSerTrpGluArgIleMetAlaTyrGluAspGlnSerIleCys 101
QY 316 ACCGTGACCCAGGACTCTCCCTCGAGGACCGCTGCTTCATCTACAGGTGAAGTT 375
Db 102 ThrAlaThrSerAspIleLysMetGluGlyAspCysPheIleTyrGluIleGlnPheHis 121
QY 376 GCGTGAAGTTCCTCCCTCCGACGCGCGCTGTATGAGAGAACACCATCGGCTGG 435
Db 122 GlyValAsnPheProAsnGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
QY 436 TCCACCGAGCGCTGTATCCCGCGAGCGCTGTGTGAGGCGGAGATCCCAAGCGCTG 495
Db 142 SerThrGluLysMetTyrValArgAspGlyValLeuLysGlyAspValAsnMetAlaLeu 161
QY 496 AAGCTGAAGGACGCGCGCTACTACTGTGAGTTCAGTCCATCTACATCCCAAG 555
Db 162 LeuLeuGluGlyGlyHisTyrArgCysAspPheArgSerThrTyrLysAlaLysLys 181
QY 556 CCGGTGACGTCGCGCGCTACTACTAGTGGACCAAGTGGACATCATCTCCCAAG 615
Db 182 ArgValGlnLeuProAspTyrHisPheValAspHisArgIleGluIleLeuSerHisAsp 201
QY 616 GAGGACTACACATCGTGGAGCAGTACGAGCGCGCGCGCGCGCGCGCGCGCTG 669
Db 202 AsnAspTyrAsnThrValLysLeuSerGluAspAlaGluAlaArgTyrSerMet 219

RESULT 6
Q95UA7 PRELIMINARY; PRT; 225 AA.
ID Q95UA7
AC Q95UA7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Cyan fluorescent protein (fragment).
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Falkowski P.G., Sun Y.;
RT "Montastraea cavernosa fluorescent protein.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY056460; AAL17905.1; -
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR000786; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
FT NON_TER 225
SQ SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;

Alignment Scores:
Pred. No.: 6.1e-42 Length: 225
Score: 675.00 Matches: 122
Percent Similarity: 73.49% Conservative: 36
Best Local Similarity: 56.74% Mismatches: 57

```


Db 202 SerHisAspLysAspTyrAsnLysValLysLeuTyrGluHisAlaGluAlaHisSerGly 221
QY 667 CTG 669
Db 222 Leu 222
RESULT 11
Q7ZOW5 PRELIMINARY; PRT; 225 AA.
ID Q7ZOW5
AC Q7ZOW5; 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Cyan fluorescent protein.
OS Montastrea cavernosa (Great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastrea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mc5;
RX MEDLINE=22689801; PubMed=12777529;
RA Kelmanson I.V.; Matz M.V.;
RT "Molecular Basis and Evolutionary Origins of Color Diversity in Great
Star Coral Montastrea cavernosa (Scleractinia: Faviida).";
RL Mol. Biol. Evol. 20:1125-1133(2003).
DR EMBL; AY181556; AAO61602.1;
SQ SEQUENCE 225 AA; 25843 MW; 13708587B7D93E35 CRC64;

Alignment Scores:
Pred. No.: 1,02e-41 Length: 225
Score: 672.00 Matches: 121
Percent Similarity: 73.02% Conservative: 36
Best Local Similarity: 56.28% Mismatches: 58
Query Match: 54.32% Indels: 0
DB: Gaps: 0

US-10-081-864-14 (1-678) x Q7ZOW5 (1-225)

QY 16 AACGTATCATCCGAGTTTCATCGCTTCAAGTGGCATGAGGGCACCGTGAACGGCCAC 75
Db 2 SerValIleLysSerValMetLysIleLysLeuHisMetAspGlyIleValAsnGlyHis 21
QY 76 GAGTTCGAGATCGAGGGGAGGGCGCGCCCTACGAGGGCCACACACCGTGAAG 135
Db 22 LysPheMetIleThrGlyGluGlyLysProPheGluGlyThrHisThrIle 41
QY 136 TTGAAGGTGACCAAGGGCGGCGCCCTGCTCGCTGGGACATCCTGTCCTCCCGAGTTC 195
Db 42 LeuLysValLysGluGlyGlyProLeuProPheAlaTyrAspIleLeuThrThrAlaPhe 61
QY 196 CAGTACGGCTCCAGGTGTACGTGAAGCACCCCGCGGACATCCCGACTACAGAGCTG 255
Db 62 GlnTyrGlyAsnArgValPheThrLysTyrProLysAspIleProAspTyrPheLysGln 81
QY 256 TCCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAACTTCGAGGACGGCGCGTGGCG 315
Db 82 SerPheProGluGlyTyrSerTrpGluArgSerMetThrPheGluAspGlnGlyValCys 101
QY 316 ACCGTGACCCAGGACTCCTCCTCGAGGACGGCTTCACTCATCTCAAGTGAAGTTCATC 375
Db 102 ThrValThrSerAspIleLysLeuGluGlyAspCysPhePheTyrGluIleArgPheTyr 121
QY 376 GCGGTGAACCTCCCTCCGACCGCCCTGATGCGAGAGAACCATGGGTGGGAGGCC 435
Db 122 GlyValAsnPheProSerSerGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
QY 436 TCCACGAGCGGCTGTATCCCGCGGACCGCGTGTGAGAGGGGAGATCCACAGGCCCTG 495
Db 142 SerThrGluAsnMetTyrValArgAspGlyValLeuLeuGlyAspValSerArgThrLeu 161
QY 496 AAGCTGAAGGACGGGCGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGCGCCAGAG 555

Q7ZOW8 PRELIMINARY; PRT; 227 AA.
ID Q7ZOW8
AC Q7ZOW8; 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
OS Montastrea cavernosa (Great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastrea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mc2;
RX MEDLINE=22689801; PubMed=12777529;
RA Kelmanson I.V.; Matz M.V.;
RT "Molecular Basis and Evolutionary Origins of Color Diversity in Great
Star Coral Montastrea cavernosa (Scleractinia: Faviida).";
RL Mol. Biol. Evol. 20:1125-1133(2003).
DR EMBL; AY181553; AAO61599.1;
SQ SEQUENCE 227 AA; 26017 MW; 5E312C54EA47F589 CRC64;

Alignment Scores:
Pred. No.: 9,38e-42 Length: 227
Score: 672.50 Matches: 120
Percent Similarity: 74.21% Conservative: 44
Best Local Similarity: 54.30% Mismatches: 54
Query Match: 54.30% Indels: 3
DB: Gaps: 1

US-10-081-864-14 (1-678) x Q7ZOW8 (1-227)

QY 16 AACGTATCATCCGAGTTTCATCGCTTCAAGTGGCATGAGGGCACCGTGAACGGCCAC 75
Db 2 SerValIleLysProAspMetLysIleLysLeuArgMetGluGlyAlaValAsnGlyHis 21
QY 76 GAGTTCGAGATCGAGGGGAGGGCGCGCCCTACGAGGGCCACACACCGTGAAG 135
Db 22 LysPheValIleGluGlyAspGlyLysGlyLysProPheGluGlyThrGlnSerMetAsp 41
QY 136 TTGAAGGTGACCAAGGGCGGCGCCCTGCTCGCTGGGACATCCTGTCCTCCCGAGTTC 195
Db 42 LeuThrValLysGluGlyAlaProLeuProPheAlaTyrAspIleLeuThrThrValPhe 61
QY 196 CAGTACGGCTCCAGGTGTACGTGAAGCACCCCGCGGACATCCCGACTACAGAGCTG 255
Db 62 AspTyrGlyAsnArgValPheAlaLysTyrProGlnAspIleProAspTyrPheLysGln 81
QY 256 TCCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATGAACTTCGAGGACGGCGCGTGGCG 315
Db 82 ThrPheProGluGlyTyrSerTrpGluArgSerMetThrTyrGluAspGlnGlyIleCys 101
QY 316 ACCGTGACCCAGGACTCCTCCCTGCGAG-----GACGGCTGCTTCACTTCAAGGTG 366
Db 102 ValAlaThrAsnAspIleThrLeuMetLysGlyValAspAspCysPheValTyrLysIle 121
QY 367 AAGTTCATCGCGTGAACCTTCCCTCCGACGGCCCGTGTATGCGAGAGAACCATGGGC 426
Db 122 ArgPheAspGlyValAsnPheProAlaAsnGlyProValMetGlnLysLysThrLeuLys 141
QY 427 TGGAGGCTCCACCGAGCGCTGTACCCCGCGGACGGCTGTGAGGGCGGAGATCCAC 486
Db 142 TrpGluProSerThrGluLysMetTyrValArgAspGlyValLeuLysGlyAspValAsn 161
QY 487 AAGGCTTGAAGTGAAGGAGCGGCGCCACTACCTGCTGGAGTTCAAGTCCATCTACATG 546
Db 162 MetAlaLeuLeuGluGlyGlyIleTyrArgCysAspPheLysThrThrTyrLys 181
QY 547 GCCAAGAGCCGTCAGCTCCCGGCTACTACTAGTGGACACCAAGCTGGACATCAC 606
Db 182 AlalysLysPheValGlnLeuProAspTyrHisPheValAspHisArgIleGluIleLeu 201
QY 607 TCCACACAGGAGACTACCATCTGAGGAGCAGTACGAGGCGCACCGGAGGGCGGCCAC 666

Db	162	LeuLeuGluGlyAspIysHisArgCysAsnPheArgSerThrTyrArgAlaLysLys	181
Qy	556	CCGTCGAGTCGCCGCGCTACTACTGTGGACACCAAGCTGGACATCACTCCCAAC	615
Db	182	GlyValValLeuProGluTyrHisPheValAspHisArgIleGluIleLeuSerHisAsp	201
Qy	616	GAGGACTACACCATCGTGGACGAGTACGACGCCACCGAGGCCGCG	660
Db	202	LysAspTyrAsnThrValGluValTyrGluAsnAlaValAlaArg	216

RESULT 12			
Q7Z0W7			
ID	Q7Z0W7	PRELIMINARY;	PRT; 234 AA.
AC	Q7Z0W7;		
DT	01-OCT-2003	(TtEMBLrel. 25, Created)	
DT	01-OCT-2003	(TtEMBLrel. 25, Last sequence update)	
DT	01-OCT-2003	(TtEMBLrel. 25, Last annotation update)	
DE	Green fluorescent protein.		
OS	Montastraea cavernosa (great star coral).		
OC	Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;		
OC	Favina; Favidae; Montastraea.		
OX	NCBI_TaxID=63558;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=mc3;		
RX	MEDLINE=22689801; PubMed=12777529;		
RA	Kalmanson I.V.; Matz M.V.;		
RT	Molecular Basis and Evolutionary Origins of Color Diversity in Great		
RI	Star Coral Montastraea cavernosa (Scleractinia: Favida).";		
RI	Mol. Biol. Evol. 20:1125-1133(2003).		
DR	EMBL; AV181554; AA061600.1; -.		
SQ	SEQUENCE 234 AA; 26813 MW; A7BBDE9F79CA8A70 CRC64;		

Align. No.:	3,71e-41	Length:	234
Score:	664.50	Matches:	118
Percent Similarity:	73.45%	Conservative:	48
Best Local Similarity:	52.21%	Mismatches:	57
Query Match:	53.72%	Indels:	3
DB:	5	Gaps:	1

US-10-081-864-14	(1-678)	x	Q7Z0W7	(1-234)
------------------	---------	---	--------	---------

```

Db      144 ArgLysThrLeuLysTrpGluProSerThrGluIleMetTy rValArgAspGlyValLeu 163
QY      472 AAGGGCGAGATCCACAAGGCCCTGAAGCTGAAGACGGCGGCCTACTACCTGGTGGAGTTC 531
Db      164 LysGlyASPvalAsnMetAlauenuenuuGluGlyGlyHisTy rArgCysAspPhe 183
QY      532 AAGTCCATCATCATGCCCAAGAAGCCGTGAGCTGCCCGCTACTACTACTACGTGGACACC 591
Db      184 LysThrThrTy rLysAlaLysLysValValArgLeuProAspTy rHisPheValAspHis 203
QY      592 AAGCTGGACATCACCTCCCACAAGAGAGACTACACCATCTGGAGCAGTAGTACGAGCCACC 651
Db      204 ArgileGluIleValserHisAspLysAspTy rAsnLysValLysLeuHisGluHisAla 223
QY      652 GAGGGCGCCACCACCTG 669
Db      224 GluAlaArgHisGlyLeu 229

RESULT 13
Q8TSF2
ID Q8TSF2 PRELIMINARY; PRT; 234 AA.
AC Q8TSF2; AC
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE McavFP 6
OS Montastrea cavernosa (great star coral).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxId=63558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21927629; PubMed=11929996;
RA Labas Y.A.; Gurskaya N.G.; Yanushevich Y.G., Fradkov A.F.,
RT Lukyanov K.A.; Lukyanov S.A.; Matz M.V.;
RL "Diversity and evolution of the green fluorescent protein family.";
DR ENBL; AY037769; AAK71335.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 234 AA; 26743 MW; C62C39ECB063A24D CRC64;

Alignment Scores:
Pred. No.: 4,4e-41 Length: 234
Score: 663,50 Matches: 117
Percent Similarity: 73,45% Conservative: 49
Best Local similarity: 51,77% Mismatches: 57
Query Match: 53,64% Indels: 3
DB: 5 Gaps: 1

US-10-081-864-14 (1-678) x Q8TSF2 (1-234)

```

[illegible]

QY 241 GACTACAGAGAGCTGCTCCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATCAACTTCGAG 300
 Db |||||
 QY 84 AspTyrPheLysGlnThrPheProGluGlyTyrPheTrpGluArgSerMetThrTyrGlu 103
 Db |||||
 QY 301 GAGCGCGGCTGGGACCGTGAACGACGAGTCTCTCCCTGCGAGCGGC-----TGC 351
 Db |||||
 QY 104 AspGlnGlyLeuLysThrAsnAspIleThrMetMetGluGlyValAspAspCys 123
 Db |||||
 QY 352 TTCATCTACAGTCAAGTTCATCGCGCTGAATTCCTCCGAGCGCGCGCTGATGCGAG 411
 Db |||||
 QY 124 PheAlaTyrLysIleArgPheAspGlyValAsnPheProAlaAsnGlyProValMetGln 143
 Db |||||
 QY 412 AAGAAGACCATGGCTGGGAGCGCTCCACGAGCGCGCTGTACCCCGCGAGCGGCTGCTG 471
 Db |||||
 QY 144 A-glyLysThrLeuLysTrpGluProSerThrGluIleMetTyrAlaArgAspGlyValLeu 163
 Db |||||
 QY 472 AGGCGGAGATCCACAGCGCTGAAGTGAAGGAGCGCGGCTACTACTGCTGGAGTTC 531
 Db |||||
 QY 164 LysGlyAspValAsnMetAlaLeuLeuLeuGluGlyGlyHisTyrArgCysAspPhe 183
 Db |||||
 QY 532 AAGTCCATCTACATGCCAAGAGCGCTGCGAGCTGCCCGGCTACTACTAGTGGACACC 591
 Db |||||
 QY 184 LysThrTyrTyrLysAlaLysLysValValArgLeuProAspTyrHisPheValAspHis 203
 Db |||||
 QY 592 AAGTGGACATCACTCCACAGAGGACTACACATCGTGGAGCAGTACGAGCGCACC 651
 Db |||||
 QY 204 A-glyLeuIleValSerHisAspLysAspTyrAsnLysValLysLeuHisGluHisala 223
 Db |||||
 QY 652 GAGGCGCGCCACCACCTG 669
 Db |||||
 QY 224 GluAlaArgHisGlyLeu 229
 Db |||||
 RESULT 14
 Q8T6U0 PRELIMINARY; PRT; 236 AA.
 AC Q8T6U0;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Green fluorescent protein.
 OS Dendronephthya sp. SSAL-2002.
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
 OC Nephthidae; Dendronephthya.
 OC NCBI_TaxID=191210;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21927629; PubMed=11929996;
 RA Labas Y.A., Gurskaya N.G., Yanushovich Y.G., Fradkov A.F.,
 RA Lukyanov K.A., Lukyanov S.A., Matz M.V.;
 RT "Diversity and evolution of the green fluorescent protein family";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
 DR EMBL; AF420591; AAM10825.1; -
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP like.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP; 1.
 DR PRODOM; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 236 AA; 26840 MW; CE1707CF9334A90 CRC64;
 Alignment Scores:
 Pred. No.: 1.34e-40 Length: 236
 Score: 657.00 Matches: 119
 Percent Similarity: 72.69% Conservative: 38
 Best Local Similarity: 55.09% Mismatches: 59
 Query Match: 53.11% Indels: 0
 DB: 5 Gaps: 0
 US-10-081-864-14 (1-678) x Q8T6U0 (1-236)
 QY 16 AACGTATCAACCGATTCATCGCTTCAAGTGGGAGCGCGTGAACCGGAC 75
 Db |||||
 QY 2 AsnLeuIleLysGluAspMetArgValLysValHisMetGluGlyAsnValAsnGlyHis 21
 Db |||||

QY 76 GAGTTCAGATCGAGCGGCGAGGCGCGCCCTACGAGGGCCACACACCGTGAAG 135
 Db |||||
 QY 22 AlaPheValIleGluGlyLysGlyArgProTyrGluGlyThrGlnThrLeuAsn 41
 Db |||||
 QY 136 TTCAAGTGTACCAAGGCGCGCCCTTCCCTCGCTGGGACATCTGCTCCCGCCAGTTTC 195
 Db |||||
 QY 42 LeuThrValLysGluGlyAlaProLeuProPheSerTyrAspIleLeuThrAlaLeu 61
 Db |||||
 QY 196 CAGTACGGTCCCAAGGTGACGAGACCGCGCGCGACATCCCGGACGAGAGCTG 255
 Db |||||
 QY 62 HisTyrGlyAsnArgValPheThrGluTyrProAlaAspIleThrAspTyrPheLysGln 81
 Db |||||
 QY 256 TCCTTCCCGAGGCTTCAAGTGGGAGCGGTGATGAATTCGAGACGCGCGCGTGGCG 315
 Db |||||
 QY 82 SerPheProGluGlyTyrSerTrpGluArgThrMetThrTyrGluAspLysGlyIleCys 101
 Db |||||
 QY 316 ACCGTGACCCAGGACTCTCTCCCTGCGAGGCGGTCTTCAATCAAGGTGAAGTTCATC 375
 Db |||||
 QY 102 ThrIleArgSerAspIleSerLeuGluGlyAspCysPhePheGlnAsnIleArgPheAsn 121
 Db |||||
 QY 376 GGCGTGAATCTCCCTCCGAGCGCGCGCTGATCAGAGAGAGACCATGGCTGGAGGCC 435
 Db |||||
 QY 122 GlyMetAsnPheProAsnGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
 Db |||||
 QY 436 TCCACCGAGCGCTGTACCCCGCGAGCGGTGTGAAGGCGGAGATCCACAGGCCCTG 495
 Db |||||
 QY 142 SerThrGluLysLeuHisValArgAspGlyLeuLeuValGlyAsnIleAsnMetAlaLeu 161
 Db |||||
 QY 496 AAGCTGAAGGAGCGCGCGCTACTCTAGTGGACACCAAGGCGGAGATCCACCGCAAC 555
 Db |||||
 QY 162 LeuLeuGluGlyGlyHisTyrLeuCysAspPheLysThrThrTyrLysAlaLysLys 181
 Db |||||
 QY 556 CCGCTGAGCTGCCCGCGCTACTACTAGTGGACACCAAGGCGGAGATCCACCGCAAC 615
 Db |||||
 QY 182 ValValGlnLeuProAspTyrHisPheValAspHisArgIleGluIleLeuSerAsnAsp 201
 Db |||||
 QY 616 GAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCCCGCCAC 663
 Db |||||
 QY 202 SerAspTyrAsnLysValLysLeuTyrGluHisGlyValAlaArgTyr 217
 Db |||||
 RESULT 15
 Q9U6Y3 PRELIMINARY; PRT; 266 AA.
 AC Q9U6Y3;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Fluorescent protein FP484.
 OS Clavularia sp.
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
 OC Clavulariidae; Clavularia.
 OC NCBI_TaxID=86521;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99436614; PubMed=10504696;
 RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
 RA Markelov M.L., Lukyanov S.A.;
 RT "Fluorescent proteins from nonbioluminescent Anthozoa species";
 RL Nat. Biotechnol. 17:969-973(1999).
 DR EMBL; AF168424; AF03374.1; -
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP like.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP; 1.
 DR PRODOM; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 266 AA; 30450 MW; B4E97406E2708854 CRC64;
 Alignment Scores:
 Pred. No.: 1.93e-40 Length: 266
 Score: 655.00 Matches: 119
 Percent Similarity: 73.27% Conservative: 40

```

Best Local Similarity: 54.84% Mismatches: 58
Query Match: 52.95% Indels: 0
DB: 5 Gaps: 0

US-10-081-864-14 (1-678) x Q9U6Y3 (1-266)

QY 19 GTCATCACCGAGTTTCATCGCTTCAAGTGGCGCATGAGGGCACCGTGAAGCCGACGAG 78
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 VALLIElyseProAspMetClyleLysLeuLysMetGluGlyAsnValAsnGlyHisAla 64
QY 79 TTCGAGATCAGGGGGAGGGGGAGGGCGGCCCTTACGAGGGCCACACACCGTGAAGTTG 138
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
65 PheValIleGluGlyGluGlyLysProTyrAspGlyThrHisThrLeuAsnLeu 84
QY 139 AAGTGTACCAAGGGCGGGCCCTCGCTTCCCTGGGACATCTGTCCCCCGCATTCGAG 198
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 GluValLysGluGlyAlaProLeuProPheSerTyrAspIleLeuSerAsnAlaPheGln 104
QY 199 TAGCGTCCAAGGTGTAGTGAAGCACCCCGCCGACATCCCGACATACAAAGAGCTGCC 258
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 TyrGlyAsnArgAlaLeuThrLysTyrProAspAspIleAlaAspTyrPheLysGlnSer 124
QY 259 TTCCCGGAGGCTTCAAGTGGAGCGGTGATGAACCTTCAGAACCGCGCGCTGGCGACC 318
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 PheProGluGlyTyrSerTrpGluArgThrMetThrPheGluAspLysGlyIleValLys 144
QY 319 GTGACCCAGGACTCTCTCCCTCAGACGGCTGCTTCATCTACAAGGTGAAGTTTCATCGC 378
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145 ValLysSerAspIleSerMetGluGluAspSerPheIleTyrGluIleArgPheAspGly 164
QY 379 GTGAACCTCCCGTCGACGGCGCGCTGATGACAGAGACACATGGCTGGGAGGCGCTCC 438
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 MetAsnPheProAsnGlyProValMetGlnLysThrLeuLysTrpGluProSer 184
QY 439 ACCGAGCGCTGTATACCCCGGACGGCTGCTGAAGGGCGAGATCCACAAGGCCCTGAAG 498
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 ThrGluIleMetTyrValArgAspGlyValLeuValGlyAspIleSerHisSerLeuLeu 204
QY 499 CTGAAGGACGGGGCCACTACTGTTGGTAGTTCAAGTTCATCTACATGCCCAAGAGCCC 558
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 LeuGluGlyGlyGlyHisTyrArgCysAspPheLysSerIleTyrIleAlaLysVal 224
QY 559 GTGACGCTGCCCGGCTACTACTAGTCGACACCAAGCTGGACATCACCTCCACACGAG 618
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 ValLysLeuProAspTyrHisPheValAspHisArgIleGluIleLeuAsnHisAspLys 244
QY 619 GACTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCGCCACCAACCTG 669
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 AspTyrAsnLysValThrLeuTyrGluAsnAlaValAlaArgTyrSerLeu 261

```

Search completed: July 29, 2004, 14:36:14
Job time : 44.3863 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 08:30:38 ; Search time 470.142 Seconds
(without alignments)
6388.441 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 707
Sequence: 1 ggtacgcgtctcgtctgtac.....ctggccacactgaagttt 707

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	707	100.0	707	6	AAD46286 Anemonia
2	694.4	98.2	1398	7	ACA61028 DNA encod
3	691.2	97.8	699	6	AAD46287 Anemonia
4	689.6	97.5	1404	7	ACA61027 DNA encod
5	676.8	95.7	696	3	AAA50885 A. sulcat
6	440.2	62.3	696	4	AAD03616 Anemonia
7	440.2	62.3	696	6	AAD46279 Anemonia
8	440	62.2	696	7	AAA50884 Anemonia
9	436.8	61.8	757	7	ABA00799 Wild type
10	436.8	61.8	935	3	AAA50883 Anemonia
11	435.2	61.6	767	7	ABA00800 Kindling
12	433.6	61.3	767	7	ABA00801 Kindling
13	417.8	59.1	687	6	ABL41173 H. crisp
14	417.8	59.1	687	6	ABL41172 H. crisp
15	417.8	59.1	955	6	AAI69492 A. sulcat
16	417.8	59.1	1396	6	ABL41174 Cr-449-ta
17	417.8	59.1	1396	7	ACA61024 DNA encod
18	417.8	59.1	1424	6	ABL41175 Cr-449-ta
19	417.8	59.1	1424	7	ACA61025 DNA encod
20	315	44.6	675	6	AAD46281 Discosoma
21	313.4	44.3	678	6	AAD46282 Discosoma
22	313.4	44.3	678	6	AAD28209 Discosoma
23	313.4	44.3	681	9	ADC24130 Discosoma

24	312.6	44.2	678	9	ADC24132	Adc24132 Discosoma
25	310.2	43.9	678	6	AAD46278	Aad46278 Discosoma
26	310.2	43.9	678	6	AAD28207	Aad28207 Discosoma
27	310.2	43.9	678	9	ADC24128	Adc24128 Discosoma
28	310.2	43.9	695	3	AAA48743	Aaa48743 Humanised
29	308.6	43.6	678	6	AAD28208	Aad28208 Discosoma
30	308.6	43.6	681	9	ADC24134	Adc24134 Discosoma
31	308.6	43.6	4692	6	AAI47954	Aal47954 Modified
32	308.6	43.6	4692	7	ACC44640	Acc44640 Vector pD
33	308.6	43.6	6893	9	ADE24111	Ade24111 Proviral
34	308.6	43.6	7910	4	AAD09979	Aad09979 pBit(dHSP
35	308.6	43.6	9320	6	ABS56664	Ab56664 Plasmid p
36	307	43.4	723	7	ABZ22476	Abz22476 Mammalian
37	307	43.4	5436	4	AAD10003	Aad10003 Plasmid p
38	299.2	42.3	6985	7	ACA55359	Ac55359 Transform
39	286.2	40.5	681	4	AAH47656	Aah47656 Anthozoan
40	284	40.2	835	6	ABA00244	Ab00244 C. Gigant
41	282.4	39.9	681	6	ABA00245	Ab00245 C. Gigant
42	271.8	38.4	678	3	AAA48750	Aaa48750 Humanised
43	271.8	38.4	678	3	AAA48748	Aaa48748 Humanised
44	271.8	38.4	678	6	AAD46280	Aad46280 Discosoma
45	270.8	38.3	760	7	ABA00806	Ab00806 Multiple

ALIGNMENTS

RESULT 1
AAD46286
ID AAD46286 standard; DNA; 707 BP.
XX
AC AAD46286;
XX
DT 27-DEC-2002 (first entry)
XX
DE Anemonia sulcata asFP595 non-aggregating mutant DNA, FP7-NA.
XX
KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunscreen; second messenger detector; asFP595 protein; NFP-7; gene;
mutant; ds.
XX
OS Anemonia sulcata.
XX
OS Synthetic.
XX
FH Key
CDS Location/Qualifiers
FT 7..702
FT /*tag= a
/product= "asFP595 non-aggregating mutant protein #1"
/note= "No start codon"
/partial
WO200268459-A2.
06-SEP-2002.
20-FEB-2002; 2002WO-US005749.
21-FEB-2001; 2001US-0270983P.
04-DEC-2001; 2001US-00006922.
(CLON-) CLONTECH LAB INC.
Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
WPI; 2002-691654/74.
P-PSDB; AAE28641.

New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant of an aggregating Chidarian chromo- or fluorescent protein or mutant for analyte detection assays or fluorescence activated cell sorting

```
PT applications.
XX Claim 4; Page 77; 80pp; English.
XX The invention relates to nucleic acid molecules encoding non-aggregating
XX chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
XX useful in analyte detection assays, as colouring agents, as markers in
XX recombinant DNA applications, as sunscreens or filters, in fluorescence
XX resonance energy transfer (FRET) applications, as biosensors in
XX prokaryotic and eukaryotic cells, in screening assays, as second
XX messenger detectors, in fluorescence activated cell sorting applications,
XX in protease cleavage assays or as fluorescent timers. The present
XX sequence is Anemonia sulcata asFP595 non-aggregating mutant DNA of the
XX invention
XX
SQ Sequence 707 BP; 137 A; 264 C; 202 G; 104 T; 0 U; 0 Other;

Query Match 100.0%; Score 707; DB 5; Length 707;
Best Local Similarity 100.0%; Pred. No. 1.2e-108;
Matches 707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATCGGCTCCCTGCTGACCGAGACCATGCTTTCAGGACCACTCCATCGAGGGCACCCTG 60
DB 1 GGATCGGCTCCCTGCTGACCGAGACCATGCTTTCAGGACCACTCCATCGAGGGCACCCTG 60

QY 61 AACGGCCACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGCACCCTAG 120
DB 61 AACGGCCACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGCACCCTAG 120

QY 121 GAGATGAAGATCGAGGTGATCGAGGGGCGCCCTGCTGCCCTTCCACATCTCTGTCC 180
DB 121 GAGATGAAGATCGAGGTGATCGAGGGGCGCCCTGCTGCCCTTCCACATCTCTGTCC 180

QY 181 ACTCTCTGCATGTACGGCTTCAAGGCTTTCATCAAGTACGTGTCGGGCATCCCGCTAC 240
DB 181 ACTCTCTGCATGTACGGCTTCAAGGCTTTCATCAAGTACGTGTCGGGCATCCCGCTAC 240

QY 241 TTCAGCAGTCCCTCCCGAGGGCTTCACTGGGAGCGCACCACTACGAGGACGGC 300
DB 241 TTCAGCAGTCCCTCCCGAGGGCTTCACTGGGAGCGCACCACTACGAGGACGGC 300

QY 301 GGCTTCTGTACCGCCCAACAGGACACCTCCCTGGAACGGGACCTGCTGTGTACAAGGTG 360
DB 301 GGCTTCTGTACCGCCCAACAGGACACCTCCCTGGAACGGGACCTGCTGTGTACAAGGTG 360

QY 361 AAGATCTGGGCAACACTTCCCGCGGACCGCCCGTGTATGAGAACAGGGCCGGCGG 420
DB 361 AAGATCTGGGCAACACTTCCCGCGGACCGCCCGTGTATGAGAACAGGGCCGGCGG 420

QY 421 TGGAGAGCCCTCCACCGAGATGCTGTACAGGTGAGCGGCTGCTGCGGCGCAGTCCCTG 480
DB 421 TGGAGAGCCCTCCACCGAGATGCTGTACAGGTGAGCGGCTGCTGCGGCGCAGTCCCTG 480

QY 481 ATGCCCTGTAGTGTCCCGCGGCTGCGCCTGACCTGACCTGACCTGACACCACTACCGC 540
DB 481 ATGCCCTGTAGTGTCCCGCGGCTGCGCCTGACCTGACCTGACCTGACACCACTACCGC 540

QY 541 TCCAAGAGCCGCTCCGCGCTGAGATGCGCGGCTTCCACCTCGAGGACCAACCGCATC 600
DB 541 TCCAAGAGCCGCTCCGCGCTGAGATGCGCGGCTTCCACCTCGAGGACCAACCGCATC 600

QY 601 GAGATCTGGAGGAGGTGGAGAGGGCAAGTGTACAGCAGTACGAGGCGCGCGTGGGC 660
DB 601 GAGATCTGGAGGAGGTGGAGAGGGCAAGTGTACAGCAGTACGAGGCGCGCGTGGGC 660

QY 661 CGCTACTGCGACGCGCGCCCTTCCAGCTGGGCGACACTGAAGCTT 707
DB 661 CGCTACTGCGACGCGCGCCCTTCCAGCTGGGCGACACTGAAGCTT 707

RESULT 2
ACA61028
ID ACA61028 standard; DNA; 1398 BP.
```

```
XX ACA61028;
XX 09-JUL-2003 (first entry)
XX DNA encoding chromoprotein AsRed-35-5D-tandem.
XX Chromo/fluorescent domain; labeled fusion protein;
XX site-specific gene modification; chromoprotein; colouring agent;
XX food composition; pharmaceutical; cosmetic; AsRed-35-5D-tandem; gene; ds.
XX Anthozoa.
XX
FH Key Location/Qualifiers
FT CDS 1..1398
FT /tag= a
FT /product= "AsRed-35-5D-tandem"
FT
XX WO2003031590-A2.
XX 17-APR-2003.
XX 10-OCT-2002; 2002WO-US032560.
XX 12-OCT-2001; 2001US-00976673.
XX 11-FEB-2002; 2002US-0356225P.
XX 22-MAY-2002; 2002US-0383336P.
XX (CLON-) CLONTECH LAB INC.
XX Lukanov SA;
XX WPI; 2003-381709/36.
XX P-PSDB; ABU09925.
XX
PT New nucleic acid encoding polypeptide products having at least two linked
PT chromo/fluorescent domains, useful for generating transgenic plants or
PT animals or site-specific gene modifications in cell lines.
XX Disclosure; Fig 5; 68pp; English.
XX
XX The invention describes a nucleic acid encoding a polypeptide product
XX comprising a first and a second chromo/fluorescent domain, optionally
XX joined by a linking domain. The first and second chromo/fluorescent
XX domains associate with each other under intracellular conditions so that
XX the encoded polypeptide assumes a tertiary structure. The nucleic acid
XX and the protein are useful in producing labeled fusion proteins that have
XX a precise and predictable signal to fusion partner ratio. The nucleic
XX acid may also be used in generating transgenic, non-human plants or
XX animals or site-specific gene modifications in cell lines. The
XX chromoproteins may be used as colouring agents, as a food composition, in
XX pharmaceuticals or cosmetics, as labels in analyte detection assays or as
XX selectable markers in recombinant DNA applications. This sequence encodes
XX chromo/fluorescent domain fusion protein AsRed-35-5D-tandem
XX
SQ Sequence 1398 BP; 270 A; 526 C; 400 G; 202 T; 0 U; 0 Other;

Query Match 98.2%; Score 694.4; DB 7; Length 1398;
Best Local Similarity 99.9%; Pred. No. 1.5e-106;
Matches 695; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCACTTCGAGGCGACCGTGAACGGC 66
DB 1 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCACTTCGAGGCGACCGTGAACGGC 60

QY 67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTTCGAGGGCACCCAGAGATG 126
DB 61 CACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTTCGAGGGCACCCAGAGATG 120

QY 127 AAGATCGAGGTGATCGAGGGGCGGCCCTTGCCTTCGCTTCCACATCTCTGTCCACCTCC 186
DB 121 AAGATCGAGGTGATCGAGGGGCGGCCCTTGCCTTCCACATCTCTGTCCACCTCC 180
```


QY 187 TGCATGATCGGCTCCAAAGGCTTCATCAAGTACGTGTCGGGCATCCCGACTACTTCAAG 246
DB 181 TGCATGATCGGCTCCAAAGGCTTCATCAAGTACGTGTCGGGCATCCCGACTACTTCAAG 240
QY 247 CAGTCCCTCCCGAGGGCTTCACTGGAGGCGACACCACTAGAGGACGGCGGCTTC 305
DB 241 CAGTCCCTCCCGAGGGCTTCACTGGAGGCGACACCACTAGAGGACGGCGGCTTC 300
QY 307 CTGACCGGCCACAGGACACTCCCTGACGCGGACTGCTTGGTGTACAAGGTGAAGATC 366
DB 301 CTGACCGGCCACAGGACACTCCCTGACGCGGACTGCTTGGTGTACAAGGTGAAGATC 360
QY 367 CTGGGCAACAATTCCTCCCGGACAGGCGGCTTCAGTACAGAACAGGCGGCGCTGGAG 426
DB 361 CTGGGCAACAATTCCTCCCGGACAGGCGGCTTCAGTACAGAACAGGCGGCGCTGGAG 420
QY 427 CCTTCCACCGAGATCGTACAGGTGACGCGGCTGCGGCGGCTTCCCTGATGCCC 485
DB 421 CCTTCCACCGAGATCGTACAGGTGACGCGGCTGCGGCGGCTTCCCTGATGCCC 480
QY 487 CTGAGTGTCCCGGCGGTGCGCACCTGACCTGCCACCTGCACACCACTACCGTCCAAAG 546
DB 481 CTGAGTGTCCCGGCGGTGCGCACCTGACCTGCCACCTGCACACCACTACCGTCCAAAG 540
QY 547 AAGCCCGCTCGCCCTGAGATGCGCGGCTTCCACTTCGAGGACCAAGCATCGAGATC 606
DB 541 AAGCCCGCTCGCCCTGAGATGCGCGGCTTCCACTTCGAGGACCAAGCATCGAGATC 600
QY 607 CTGGAGGAGTGGAGAGGGCAAGTGTACAGGAGTACGAGGCGGCGGCTGGCGCTTAC 666
DB 601 CTGGAGGAGTGGAGAGGGCAAGTGTACAGGAGTACGAGGCGGCGGCTGGCGCTTAC 660
QY 667 TCGGACGCGCGCCCTCCAAAGTGGGCGCAACTGA 702
DB 661 TCGGACGCGCGCCCTCCAAAGTGGGCGCAACTGA 696

RESULT 3

AAD46287
ID AAD46287 standard; DNA; 699 BP.
XX
AC AAD46287;
XX
DT 27-DEC-2002 (first entry)
XX
DE Anemonia sulcata asFP595 non-aggregating mutant DNA, Asred M35-5D.
XX
KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunsreen; second messenger detector; asFP595 protein; NFP-7; gene;
KW mutant; ds.
XX
OS Anemonia sulcata.
OS Synthetic.
XX
FH Key Location/Qualifiers
CDS 1..699
FT /*tag= a
FT /product= "asFP595 non-aggregating mutant protein"
XX
XX WO200268459-A2.
XX
PD 06-SEP-2002.
XX
XX 20-FEB-2002; 2002WO-US005749.
XX
XX 21-FEB-2001; 2001US-0270983P.
XX
XX 04-DEC-2001; 2001US-00006922.
XX
XX (CLON-) CLONTECH LAB INC.

PI Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Pradkov A;
XX WPI; 2002-691654/74.
DR P-PSDB; AAE28842.
XX
XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
XX of an aggregating Chidarian chromo- or fluorescent protein or mutant for
XX analyte detection assays or fluorescence activated cell sorting
XX applications.
XX
PS Claim 4; Fig 13; 80pp; English.
XX
CC The invention relates to nucleic acid molecules encoding non-aggregating
CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC useful in analyte detection assays, as colouring agents, as markers in
CC recombinant DNA applications, as sunsreens or filters, in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, in screening assays, as second
CC messenger detectors, in fluorescence activated cell sorting applications,
CC in protease cleavage assays or as fluorescent timers. The present
CC sequence is Anemonia sulcata asFP595 non-aggregating mutant DNA of the
CC invention. Note: This sequence is stated to be the same as that shown as
CC SEQ ID NO:23 in the sequence listing. However this sequence contains
CC additional bases at its 5' end
XX
SQ Sequence 699 BP; 137 A; 261 C; 200 G; 101 T; 0 U; 0 Other;
Query Match 97.8%; Score 691.2; DB 6; Length 699;
Best Local Similarity 99.6%; Pred. No. 5.2e-106;
Matches 693; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 7 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 66
DB 4 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 63
QY 67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGCACCCAGGAGATG 126
DB 64 CACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGCACCCAGGAGATG 123
QY 127 AAGATCGAGGTGATCGAGGGGCGGCCCTCGCTTGCCTTGCACATCTCTGCACCTCC 186
DB 124 AAGATCGAGGTGATCGAGGGGCGGCCCTCGCTTGCCTTGCACATCTCTGCACCTCC 183
QY 187 TGCATGTACGGCTCCAAAGGCGCTTCATCAAGTACGTGTCGGGCATCCCGCACTACTTCAAG 246
DB 184 TGCATGTACGGCTCCAAAGGCGCTTCATCAAGTACGTGTCGGGCATCCCGCACTACTTCAAG 243
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGGCGACCAACCACTACGAGGACGGCGGCTTC 306
DB 244 CAGTCCCTCCCGAGGGCTTCACCTGGAGGCGACCAACCACTACGAGGACGGCGGCTTC 303
QY 307 CTGACCGGCCACAGGACACCTCCCTCGAGCGGCGACTGCTGTGTGTACAAGGTGAAGATC 366
DB 304 CTGACCGGCCACAGGACACCTCCCTCGAGCGGCGACTGCTGTGTGTACAAGGTGAAGATC 363
QY 367 CTGGGCAACAATTCCTCCCGGCGAGCGGCTTCAGTACAGAACAGGCGGCGGCTGGAG 426
DB 364 CTGGGCAACAATTCCTCCCGGCGAGCGGCTTCAGTACAGAACAGGCGGCGGCTGGAG 423
QY 427 CCCTCCACCGAGATCGTGTACAGGTGAGCGGCTGCTGCGGCGGCTGCTGCTGATGCC 486
DB 424 CCCTCCACCGAGATCGTGTACAGGTGAGCGGCTGCTGCGGCGGCTGCTGCTGATGCC 483
QY 487 CTGGAGTGTCCCGGCGGCTGCGCACTGACCTGCGACCTGCGACACCACTACCGTCCAAAG 546
DB 484 CTGGAGTGTCCCGGCGGCTGCGCACTGACCTGCGACCTGCGACACCACTACCGTCCAAAG 543
QY 547 AAGCCCGCTCGCCCTGAGATGCGCGGCTTCACCTTCGAGGACCAAGCATCGAGATC 606
DB 544 AAGCCCGCTCGCCCTGAGATGCGCGGCTTCACCTTCGAGGACCAAGCATCGAGATC 603
QY 607 CTGGAGGAGTGGAGAGGGCAAGTGTGTACAAGCAGTACGAGGCGGCGGCTGGCGCTTAC 666

Db 601 ATGGAGGAGGTGGAGAGGCGCAAGTCTACAAGCAGTACGAGCGCGCTGGCGCGCTAC 696

Qy 667 TCGACGCGCGCCCTCCAAAGCTGGGCCCAACTGA 702

Db 661 TCGACGCGCGCCCTCCAAAGCTGGGCCCAACTAA 696

RESULT 6

AA003616

ID AAD03616 standard; cDNA; 696 BP.

XX AAD03616;

XX 19-JUN-2001 (first entry)

XX Anemonia sulcata chromo/fluorescent protein, asFP600 (NFP-7) cDNA.

XX Anthozoa; Chromoprotein; fluorescent protein; asFP600; NFP-7; sunscreen;

KW analyte detection assay; selectable marker; recombinant DNA application;

KW biosensor; pH indicator; invivo marker; selective filter; sea anemone;

KW SS.

XX Anemonia sulcata.

XX OS

XX Key Location/Qualifiers

CDS 1..696

FT /*tag= a

FT /product= "Chromo/fluorescent protein, asFP600 (NFP-7) "

FT /note= "CDS does not include stop codon"

FT /partial

XX WO200127150-A2.

XX PN

XX 19-APR-2001.

XX PD

XX 13-OCT-2000; 2000WO-US028477.

XX PF

XX 14-OCT-1999; 99US-00418529.

PR 15-OCT-1999; 99US-00418917.

PR 15-OCT-1999; 99US-00418922.

PR 19-NOV-1999; 99US-00443338.

PR 19-NOV-1999; 99US-00444341.

PR 09-DEC-1999; 99US-00457556.

PR 09-DEC-1999; 99US-00457899.

PR 09-DEC-1999; 99US-00458144.

PR 09-DEC-1999; 99US-00458477.

PR 10-DEC-1999; 99WO-US029405.

PR 14-JUN-2000; 2000US-0211607P.

PR 14-JUN-2000; 2000US-0211609P.

PR 14-JUN-2000; 2000US-0211626P.

PR 14-JUN-2000; 2000US-0211627P.

PR 14-JUN-2000; 2000US-0211687P.

PR 14-JUN-2000; 2000US-0211766P.

PR 14-JUN-2000; 2000US-0211880P.

PR 14-JUN-2000; 2000US-0211888P.

PR 14-JUN-2000; 2000US-0212070P.

XX XX

PA (CLON-) CLONTECH LAB INC.

XX Lukanov SA, Fradkov AF, Labas YA, Matz MV, Tersikh A;

XX WFI; 2001-266409/27.

DR P-PSDB; AAE00376.

XX XX

XX An Anthozoa chromo- or fluorescent protein (Pt) present in an environment other than its natural environment, useful as a label in analyte detection assays and as a selectable marker in recombinant DNA applications.

XX Claim 13; Fig 7; 69pp; English.

XX The present sequence is an Anemonia sulcata chromo/fluorescent protein,

asFP600 (NFP-7) cDNA. NFP-7 is present in an environment other than its natural environment and has an absorbance maximum ranging from 370nm to 770nm and more usually from 560nm to 580nm and emission maximum ranging from 395nm to 795nm and more usually from 585 to 605nm. The chromoproteins or fluorescent proteins are useful as labels in analyte detection assays, as selectable markers in recombinant DNA applications, as biosensors in prokaryotic and eukaryotic cells e.g. as pH indicator and as in vivo markers in animals. They are also useful in sunscreens and as selective filters. Note: The specification also refers to a mutant of this sequence, Mut1 with substitutions T70A and A148S. However, the position of nucleotide 'A' do not match with this wild type sequence

XX
SQ Sequence 696 BP; 195 A; 147 C; 176 G; 178 T; 0 U; 0 Other;

Query Match 62.3%; Score 440.2; DB 4; Length 696;
Best Local Similarity 77.2%; Pred. No. 2.2e-64;
Matches 535; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy 7 GCCTCCCTGTCACCGAGACATGCCCTTCAGGACACCATCGAGGCGACCGTGAACGGC 66
Db 4 GCTTCCTTTTAAAGAAAGACTATGCCCTTTAAGACGACCATTAAGGGGACGGTTAATGGC 63

Qy 67 CACTACTTCAAGTCCACCGAGGGGAGGGCAACCCCTCGAGGGCACCCAGGAGATG 126
Db 64 CACTACTTCAAGTGTACAGAAAGAGGAGGGCAACCCATTGAGGGTACGAGGAATG 123

Qy 127 AAGATCGAGGTGATCGAGGGGGGGCCCTCGCTTCCTTCACATCTGCTCCACCTCC 186
Db 124 AAGATAGAGGTGATCGAAGGAGGTCCATTGCCATTTGCCCTTCACATTTTGTCAACGAGT 183

Qy 187 TGCATGTACGGCTCCAGGGCTTCATCAAGTAGCTGTCGGGATCCCGACTACTTCAG 246
Db 184 TGTATGACGGTAGTAGAGGCTTCATCAAGTAGTGTGAGGAATTCCTGACTACTTCAAG 243

Qy 247 CAGTCCCTCCCGAGGCTTCACCTGGGAGCGCACCAACCTACGAGGACCGCGCTTC 306
Db 244 CAGTCTTTCCCTGAAGGTTTACTTGGAAAGAACCAACCACTACGAGGATGAGGCTTT 303

Qy 307 CTGACCGCCACGAGACACTCCCTGACGGGAGTGGCTGTGTACAGGTGAGATC 366
Db 304 CTTACAGCTCATCAGGACAAAGCCTAGATGAGATGTCCTGTTTACAAAGTCAAGATT 363

Qy 367 CTGGGCAACACTTCCCGCGGAGCGCCCGTATGAGAACAAAGCGCGCGCTGGGAG 426
Db 364 CTGGTAATTAATTTCTGCTGATGCCCGTGTATGAGAACAAAGAGATGGGAG 423

Qy 427 CCTTCAACGAGATCGGTGACAGGTGAGCGGCTGTCGGGGCCAGTCCCTGATGGCC 486
Db 424 CCATCCACGAGATAGTTTATGAAGTTGACGGTGTCTCGTGGACAGTCTTGTATGGCC 483

Qy 487 CTGAGTGGCCCGCGGTGCCACCTGACCTGCCACTTCGACACCACTACCGTCCAG 546
Db 484 CTTAAGTGGCCCTGGTGGTGTCTATCTGACTTGCATCTCATACTACTTACAGGTCCAA 543

Qy 547 AAGCCCGCTCCCGCTGAGATGCCCGGTTCCTCACTTCGAGAACACCGCATCGAGATC 606
Db 544 AAACCAAGCTCTCGCTTGAAGATGTCAGGATTTCAATTTGAAGATCATCGCATCGAGATA 603

Qy 607 CTGAGGAGGTGAGAGGAGCAAGTCTACAGCAGTACGAGCGCCCTGGGCGCTAC 666
Db 604 ATGAGAGAGTGTGAGAAAGCAAGTGTCTATTAACAGTACGAAGCAGTGGGCGAGTAC 663

Qy 667 TGGAGCGCCCGCCCTCCAGCTGGGCCAAC 699
Db 664 TGTGATGCTGCTCCATCCAGCTTGGACATAAC 696

RESULT 7

AAD46279

ID AAD46279 standard; DNA; 696 BP.

XX

AC AAD46279;

XX

DT 27-DEC-2002 (first entry)
DE Anemonia sulcata asFP600 (NFP-7) wild-type protein encoding DNA.
XX
KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
KW fluorescence activated cell sorting application; fluorescent timer;
KW biosensor; fluorescence resonance energy transfer application; FRET;
KW colouring agent; recombinant DNA application; analyte detection assay;
KW sunsreen; second messenger detector; asFP600 protein; NFP-7; gene; ds.
XX
OS Anemonia sulcata.
XX
FH Key Location/Qualifiers
FT CDS 1..696
FT /*tag= a
FT /product= "asFP600 wild-type protein"
FT /transl_except= (pos:202..204, aa:Thr)
FT /transl_except= (pos:427..429, aa:Ala)
FT /note= "No stop codon"
FT /partial
XX
PN W0200268459-A2.
PD
XX
PD 06-SEP-2002.
XX
PF 20-FEB-2002; 2002WO-US005749.
XX
XX 21-FEB-2001; 2001US-0270983P.
PR 04-DEC-2001; 2001US-00006922.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukanov S, Lukanov K, Yanushevich Y, Savitsky A, Pradkov A;
PI WPI; 2002-691654/74.
DR P-PSDB; AAE28834.
XX
XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
XX of an aggregating Chidarian chromo- or fluorescent protein or mutant for
XX analyte detection assays or fluorescence activated cell sorting
XX applications.
XX
XX Disclosure; Page 71; 80pp; English.
XX
XX The invention relates to nucleic acid molecules encoding non-aggregating
XX chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
XX useful in analyte detection assays, as colouring agents, as markers in
XX recombinant DNA applications, as sunscreens or filters, in fluorescence
XX resonance energy transfer (FRET) applications, as biosensors in
XX prokaryotic and eukaryotic cells, in screening assays, as second
XX messenger detectors, in fluorescence activated cell sorting applications,
XX in protease cleavage assays or as fluorescent timers. The present
XX sequence is a DNA encoding Anemonia sulcata asFP600 (NFP-7) wild-type
XX protein of the invention
SQ Sequence 696 BP; 195 A; 147 C; 176 G; 178 T; 0 U; 0 Other;

Query Match

Best Local Similarity 77.2%; Score 440.2; DB 6; Length 696;

Matches 535; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy 7 GCCTCCCTGTCACCGAGACATGCCCTTCAGGACACCATCGAGGCGACCGTGAACGGC 66

Db 4 GCTTCCTTTTAAAGAAAGACTATGCCCTTTAAGACGACCATTAAGGGGACGGTTAATGGC 63

Qy 67 CACTACTTCAAGTCCACCGGAGGGGAGGGCAACCCCTCGAGGCGACCCAGGAGATG 126

Db 64 CACTACTTCAAGTGTACAGAAAGAGGAGGGGCAACCCATTGAGGTACGAGGAATG 123

Qy 127 AAGATCGAGGTGATCGAGGGGGCGCCCTCGCTTCCTTCACATCTGCTCCACCTCC 186

Db 124 AAGATAGAGGTGATCGAAGGAGGTTCATTGCCATTTGCCCTTCACATTTTGTCAACGAGT 183

RESULT 8					
AAA50884					
ID AAA50884 standard; cDNA; 696 BP.					
XX					
AC AC					
DT 08-SEP-2000 (first entry)					
XX	Anemonia sulcata mutant fluorescent protein, Mut1, coding sequence.				
DE					
XX					
KW Green fluorescent protein; GFP; Mut1; protein localisation;					
XX fluorescence resonance energy transfer; ss.					
OS Anemonia sulcata.					
XX					
Key Location/Qualifiers					
CDS 1..696					
FT /*tag= a					
FT /partial					
FT /product= "Mut1"					
XX					
PN WO200034319-A1.					
XX					
PD 15-JUN-2000.					
XX					
PF 10-DEC-1999; 99WO-US029300.					
XX					
PR 11-DEC-1998; 98US-00210330.					
PR 09-DEC-1999; 99US-00210330.					
XX					
PA (CLON-) CLONTECH LAB INC.					
XX					
Lukyanov SA, Pradkov AF, Labas YA, Matz MV, Jiang X, Duong T;					
Pi Zhao X;					
DR WPI; 2000-423374/36.					


```
PR 11-DEC-1998; 98US-00210330.
PR 09-DEC-1999; 99US-00210330.
XX PA (CLON-) CLONTECH LAB INC.
XX PI Lukyanov SA, Fradkov AF, Labas VA, Matz MV, Jiang X, Duong T;
XX Zhao X;
XX WPI; 2000-423374/36.
XX P-PSDB; AAY99733.
XX Novel fluorescent proteins from non-bioluminescent Anemonia sulcata,
XX useful for fluorescent labeling and as markers.
XX Claim 8; Page 63-64; 71pp; English.
XX The present sequence is the coding sequence for fluorescent protein,
XX asFP600, from non-bioluminescent Anemonia sulcata. The fluorescent
XX protein (AAY99733) has applications in fluorescent labelling, as
XX fluorescent markers for gene expression and protein localisation studies,
XX and in fluorescence resonance energy transfer (FRET) reactions. The
XX coding sequence may be used as a source or primers and probes for
XX identifying related proteins
XX SQ Sequence 955 BP; 273 A; 196 C; 216 G; 270 T; 0 U; 0 Other;
Query Match 61.8%; Score 436.8; DB 3; Length 955;
Best Local Similarity 76.7%; Pred. No. 7.8e-64;
Matches 534; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
QY 7 GCCTCCCTGTCACGAGACCATGCGCTTCAGACACCATCGAGGCGACCGTGAACGCG 66
DB 72 GCCTCCCTTTTAAAGAAGACTATGCCCTTTAAGACGACCATTTGAAGGCGCGTTAATGCG 131
QY 67 CACTACTTCAAGTGACCGGAGGCGGAGGCGCAACCCCTCGAGGCGACCCGAGGAGATG 126
DB 132 CACTACTTCAAGTGATACGAAAGAGAGGCGCAACCCATTGAGGCGTACGAGGAGATG 191
QY 127 AAGATCAGGTGATCAGGCGCGCCCTGCGCTTCGCCTTCCACATCGCTGCCACTCC 186
DB 192 AAGATAGAGGTATCAGGAGGAGTCCATTGCCATTGGCTTCCACATTTGTCAACGAGT 251
QY 187 TGCATCTAGCGTCCAGGCGCTTCATCACTAGTCTGCGGCGATCCCGACTACTTCAG 246
DB 252 TGTATGACGGTATGTAAGACCTTCATCAAGTATGTGTGAGGAATTCCTGACTACTTCAG 311
QY 247 CAGTCCCTCCCGAGGCGCTTCACCTGGAGCGGCACCAACCTACGAGGCGCGGCTTC 306
DB 312 CAGTCTTCCCTGAGGTTTACTTGGGAAGAACCAACCTACGAGGATGAGGCTTT 371
QY 307 CTGACGCGCCACAGGACACCTCCCTGACGCGGACTGCGTGTGTACAGGTGAAGTC 366
DB 372 CTTACAGCTCATCAGGACACAGCTAGATGGAGATTGCTCGTTTACAGGTCAAGATT 431
QY 367 CTGGGCAACAACCTCCCGCGGCGCGCTGATGCAGAACCAAGCGCGCGCTGGGAG 426
DB 432 CTGGTATATATTTCTGCTGATGGCCCGCTGATGCAGAACCAAGCAGGAGATGGAG 491
QY 427 CCCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGCTGCGCGCCAGTCCCTGATGGCC 486
DB 492 CCAGCCACCGAGATAGTTTATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 551
QY 487 CTGGAGTGGCCCGGCGGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 546
DB 552 CTTAAGTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 611
QY 547 AAGCCCGCTCCGCGCTGAGATGCGCGCTTCCACTTTCGAGGACCAACCGCATCGAGATC 606
DB 612 AACCAGCTAGTGCCTTGAAGATGCCAGGATTTTCATTTGAAGATCATCGATCGAGATA 671
QY 607 CTGAGGAGGTGAGAGGCGCAAGTGTCAACAGCAGTACGAGCGCGCGTGGCCCTAC 666
DB 672 ATGAGGGAAGTTGAGAAAGGCAAGTGTCTATAAACAGTACAGCAGCAGTGGCGAGTAC 731
```

```
QY 667 TGGACGCGCGCCCTCCAGCTGGGCGCACCACTGA 702
DB 732 TGTGATGCTGCTCCATCCAGCTTGGACATACTAA 767
```

```
RESULT 11
ABA00800
ID ABA00800 standard; cDNA; 767 BP.
XX ABA00800;
AC ABA00800;
XX 01-APR-2003 (first entry)
DT Kindling fluorescent protein, AsFP595 A148G coding sequence.
XX Gene; kindling fluorescent protein; kindling stimulus; movement;
XX labeling; fluorescence resonance energy transfer; FRET;
XX bioluminescence resonance energy transfer; BRET; biosensor;
XX automated screening; ss.
XX Anthozoa sulcata.
XX Location/Qualifiers
FH Key 69..763
FT CDS /*tag= a
FT /product= "AsFP595 A148G"
FT mutation replace(496,C)
FT /*tag= b
XX WO200296924-A1.
XX 05-DEC-2002.
XX 24-MAY-2002; 2002WO-US016379.
XX 25-MAY-2001; 2001US-0293752P.
XX 11-OCT-2001; 2001US-0329176P.
XX (CLON-) CLONTECH LAB INC.
XX Lukyanov SA, Chudakov D, Lukyanov K;
XX WPI; 2003-156788/15.
XX P-PSDB; AAG79763.
XX Novel nucleic acid that is present in other than its natural environment
XX and that encodes kindling fluorescent protein, is useful in labeling
XX protocols, e.g. labeling proteins, organelles, cells and organisms.
XX Example; Fig 2; 96pp; English.
```

The sequences given in ABA00799-805 encode wild type and mutant kindling fluorescent proteins. The proteins go from a first substantially non-fluorescent or non-fluorescent state to a second fluorescent state upon exposure to a kindling stimulus. The kindling proteins are useful for detecting an entity such as a protein, organelle or cell in a composition such as a cell or a multicellular composition (preferably a multicellular organism), by providing the entity as an entity labeled with the kindling protein, kindling the kindling fluorescent protein label with a kindling stimulus to produce a kindled kindling fluorescent protein label, and exciting the kindled kindling fluorescent protein label with light and detecting any fluorescence from it to detect the entity. The method monitors the movement of the entity. The fluorescent proteins and the cDNA encoding them are useful in labeling protocols, e.g. labeling proteins, organelles, cells and organisms, as biological labels or markers, in protein labeling or tagging applications. The fluorescent kindling proteins are useful as detectable labels, as labels in analyte detection assays, in bioluminescence resonance energy transfer (BRET) applications, as biosensors in prokaryotic and eukaryotic cells, in applications involving the automated screening of arrays of cells expressing fluorescent reporting groups, in high through-put screening

```
CC assays, as second messenger detectors, and in fluorescent activated cell
CC sorting assays
XX
SQ Sequence 767 BP; 225 A; 162 C; 188 G; 192 T; 0 U; 0 Other;

Query Match 61.6%; Score 435.2; DB 7; Length 767;
Best Local Similarity 76.6%; Pred. No. 1.5e-63;
Matches 533; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 7 GCCTCCCTGCTGACCGAGACCATCCCTTCAGGACCACTCGAGGCGACCGTGAACGGC 66
Db 72 GCTTCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTTGAAGGACGGTTAATGCG 131
QY 67 CACTACTTCAAGTCACCGGCAAGGGGAGGGGACCCCTCGAGGCGACCCAGGAGATG 126
Db 132 CACTACTTCAAGTCACCGGCAAGGGGAGGGGACCCCTCGAGGCGACCCAGGAGATG 191
QY 127 AAGATCGAGGTGATCGAGGGGGCCCTCGCTTCGCCCTTCACATCCTGTCCACCTCC 186
Db 192 AAGATAGAGTTCATCGAGGAGTCCATTCGCCATTTGCCCTTCACATTTTGTCAACGAGT 251
QY 187 TGCATGTACGCTCCAGGCGCTTCATCAAGTACGTGTCGGGATTCGCCGACTACTTCAAG 246
Db 252 TGTATGTACGCTAGTAAAGACCTTCATCAAGTATGTGTACGAAATTCCTGACTACTTCAAG 311
QY 247 CAGTCCCTCCCGAGGCGCTTCACCTGGGAGCGCACCACTACGAGGACGGCGCTTC 306
Db 312 CAGTCTTCCCTGAAGTTTCTTGGGAAAGAACCAACCTACGAGGATGAGGCTTT 371
QY 307 CTGACCGCCCAACAGGACACCTCCCTCGACGGGACCTGCTGTGTACAAAGTCAAGATC 366
Db 372 CTTACAGTCTATCAGGACACAAAGCTTAGATGGAGATTGCCCTGTTTACAAAGTCAAGATT 431
QY 367 CTGGGCAACACTTCCCGCGAGCGCGCGGTGATGACAGAACAGCGCGCGCTGGAG 426
Db 432 CTTGGTAATATTTTCTGCTGTATGATGGCGCGGTGATGACAGAACAGGAAAGATGGAG 491
QY 427 CCCTCCACCGAGATCGTGTACAGGTGGAGCGGTGCTGCGCGGCCAGTCCCTGATGGCC 486
Db 492 CCAGGACCGAGATGTTTATGAAGTTGACGGTGTCTGCTGTCGACAGTCTTTGATGGCC 551
QY 487 CTGAGTGCCTCCCGCGGTGCGCACTGACCTGCGCACTGCGACACCTACCTCCGCTCAAG 546
Db 552 CTTAAGTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 611
QY 547 AAGCCGCGCTCCGCGCTGAAAGATCCCGGCTTCACCTTCGAGGACCAACCGCATCGAGATC 606
Db 612 AAACAGCTAGTGTCTGAGATGCCAGGATTTTATTTTGAAGTATCATCGATCGAGATA 671
QY 607 CTGAGAGGTGGAGAGGCAAGTGTACAAGCAGTACGAGGCGCGCGTGGCGCGCTAC 666
Db 672 ATGAGGAAGTTGAGAAGGCAAGTGTATAAAGCAGTACGAGCAGTGTGGGCGAGTAC 731
QY 667 TGCCAGCGCGCGCTCCAGCTGGGCGCACTGA 702
Db 732 TGTGATGCTGCTCCATCCAGCTTGGACATACATAA 767

RESULT 12
ABA00801
ID ABA00801 standard; cDNA; 767 BP.
XX
AC ABA00801;
XX
DT 01-APR-2003 (first entry)
XX
DE Kindling fluorescent protein, AsFP595 F90L A148G H203Y cDNA.
XX
KW Gene; kindling fluorescent protein; kindling stimulus; movement;
KW labeling; fluorescence resonance energy transfer; FRET;
KW bioluminescence resonance energy transfer; BRET; biosensor;
KW automated screening; ss.
XX
```

```
OS Anthozoa sulcata.
XX
FH Key Location/Qualifiers
FT CDS 59..763
FT /*tag= a
FT /product= "AsFP595 F90L A148G H203Y"
FT /replace(330,T)
FT /*tag= b
FT /replace(496,C)
FT /*tag= c
FT /replace(657..659, CAT)
FT /*tag= d
XX
XX WO200296924-A1.
PN
XX
XX 05-DEC-2002.
XX
XX 24-MAY-2002; 2002WO-US016379.
XX
XX 25-MAY-2001; 2001US-0293752P.
PR 11-OCT-2001; 2001US-0329176P.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukanov SA, Chudakov D, Lukyanov K;
XX
XX WPI: 2003-156788/15.
DR P-PSDB; AAG79764.
XX
XX Novel nucleic acid that is present in other than its natural environment
PT and that encodes kindling fluorescent protein, is useful in labeling
PT protocols, e.g. labeling proteins, organelles, cells and organisms.
XX
XX Example; Fig 3; 96pp; English.
XX
XX The sequences given in ABA00799-805 encode wild type and mutant kindling
CC fluorescent proteins. The proteins go from a first substantially non-
CC fluorescent or non-fluorescent state to a second fluorescent state upon
CC exposure to a kindling stimulus. The kindling proteins are useful for
CC detecting an entity such as a protein, organelle or cell in a composition
CC such as a cell or a multicellular composition (preferably a multicellular
CC organism), by providing the entity as an entity labeled with the kindling
CC protein, kindling the kindling fluorescent protein label with a kindling
CC stimulus to produce a kindled kindling fluorescent protein label, and
CC exciting the kindled kindling fluorescent protein label with light and
CC detecting any fluorescence from it to detect the entity. The method
CC monitors the movement of the entity. The fluorescent proteins and the
CC cDNA encoding them are useful in labeling protocols, e.g., labeling
CC proteins, organelles, cells and organisms, as biological labels or
CC markers, in protein labeling or tagging applications. The fluorescent
CC kindling proteins are useful as detectable labels, as labels in analyte
CC detection assays, in fluorescence resonance energy transfer (FRET)
CC applications, in bioluminescence resonance energy transfer (BRET)
CC applications, as biosensors in prokaryotic and eukaryotic cells, in
CC applications involving the automated screening of arrays of cells
CC expressing fluorescent reporting groups, in high throughput screening
CC assays, as second messenger detectors, and in fluorescent activated cell
CC sorting assays
XX
SQ Sequence 767 BP; 225 A; 163 C; 188 G; 191 T; 0 U; 0 Other;
```

```
Query Match 61.3%; Score 433.6; DB 7; Length 767;
Best Local Similarity 76.4%; Pred. No. 2.7e-63;
Matches 532; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 7 GCCTCCCTGCTGACCGAGACCATCCCTTCAGGACCACTCGAGGCGACCGTGAACGGC 66
Db 72 GCTTCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTTGAAGGACGGTTAATGCG 131
QY 67 CACTACTTCAAGTCACCGGCAAGGGGAGGGGACCCCTTCGAGGCGACCCAGGAGATG 126
Db 132 CACTACTTCAAGTCACCGGCAAGGGGAGGGGACCCCTTCGAGGCGACCCAGGAGATG 191
```


127	QY	127	AGATCGAGGTGATCGAGGGCGGCCCCCTTCGCCCTTCACATPCTCTCCACCTCC	186
128	DB	128		
129	DB	129		
130	DB	130		
131	DB	131		
132	DB	132	AGATAGAGGTGATCGAGGAGGTCCATGTCATTTGCCATTTGCCCTTCACATTTTGTCAACGAGT	251
133	QY	133		
134	DB	134		
135	DB	135		
136	DB	136		
137	QY	137	TGATGTACGGCTCCAAAGCGCTTCATCAAGTAGTGTCGGGATCCCGGACTACTTCAAG	246
138	DB	138		
139	DB	139		
140	DB	140		
141	DB	141		
142	QY	142	CAGTCCCTCCCGGAGGGCTTCACCTGGGAGCGCACCAACCTACGAGAGCGCGGCTTC	306
143	DB	143		
144	DB	144		
145	DB	145		
146	QY	146	CAGTCTTTTCCCTGAAGGTCTTACTTGGGAAAGAACCAACAACCTACGAGGATGGAGGCTTT	371
147	QY	147		
148	QY	148	CTGACCGCGCCACCGGACACCTCCCTGGAGCGGACTGCCTGTGTACAAAGGTGAAGATC	366
149	DB	149		
150	DB	150		
151	DB	151		
152	DB	152		
153	DB	153		
154	QY	154	CTTACAGCTCATCAGGACACAAGCCTAGATGGAGTTGCCCTCGTTTACAAAGTCAAGATT	431
155	QY	155		
156	QY	156	CTGGGCAACAACCTTCCCGCGCAGCGGCCCGTGATGCAGAAACAGCGCGCTCGGAG	426
157	DB	157		
158	DB	158		
159	DB	159		
160	DB	160		
161	QY	161	CTTGGTAATAATTTTCTGTGCTGATGGCCCGGTGATGCAGAAACAAGCAGGAAGATGGGAG	491
162	QY	162		
163	QY	163	CCCTCCACGGAGATCGTGTACGAGGTGGACGGCGTCTGCGCGGCCAGTCCCTGATGGCC	486
164	DB	164		
165	DB	165		
166	DB	166		
167	QY	167	CCAGCACCGGAGATAGTTTATTAAGATTGAOGGTCTCTGCTGGACAGTCTTTTGATGGCC	551
168	QY	168		
169	QY	169	CTGGAGTGCCTCGCGCGGTGCGCACCTGACCTGCCACCTGCACACCACTACCGCTCCAA	546
170	DB	170		
171	DB	171		
172	DB	172		
173	DB	173		
174	DB	174		
175	QY	175	CTTAAGTGCCTGTGGTGTGTCATCTGACTTGCCATCTCCATACTACTACAGTCCAA	611
176	QY	176		
177	QY	177	AAGCCGCTCGCGCTCGAAGATCGCGGCTTCCACTTCGAGACACACCGCATCGAGATC	606
178	DB	178		
179	DB	179		
180	DB	180		
181	QY	181	AAACCAAGCTAGTGCCTTGAAGATCCAGGATTTCATTTTGAAGATTACCGCATCGAGATA	671
182	QY	182		
183	DB	183		
184	DB	184		
185	QY	185	CTGGAGGAGGTGGAGAAAGCAAGTGTCTACAAGCAGTACGAGGCGCGCGTGGCCGCTAC	666
186	DB	186		
187	DB	187		
188	DB	188		
189	DB	189		
190	QY	190	ATGGAGGAAGTTGAGAAAGGAAGTGTCTATAAACAAGTACGAGCAGTGGCAGGTAC	731
191	QY	191		
192	QY	192	TGCGACGCGCGCCCTCCAAAGCTGGGCCACAACCTGA	702
193	DB	193		
194	DB	194		
195	DB	195	TGTGATGCTGTCTCCATCCAAAGCTTGGACATAACTAA	767
196	DB	196		
197	DB	197		
198	DB	198		
199	DB	199		
200	DB	200		
201	DB	201		
202	DB	202		
203	DB	203		
204	DB	204		
205	DB	205		
206	DB	206		
207	DB	207		
208	DB	208		
209	DB	209		
210	DB	210		
211	DB	211		
212	DB	212		
213	DB	213		
214	DB	214		
215	DB	215		

PR	16-JUL-2001; 2001US-0306131P.	
XX	(CLON-) CLONTECH LAB INC.	
XX	Lukyanov SA, Fradkov AF, Lukyanov KA, Gurskaya NG;	
XX	WPI; 2002-444170/47.	
XX	P-PSDB; ABB07995.	
XX	Novel nucleic acid encoding Stichodactyliden chromoprotein and its	
PT	fluorescent mutant useful as coloring agent, labels in analyte detection	
PT	assays, markers in recombinant DNA applications and filters in	
PT	sunscreens.	
XX	Claim 5; Fig 10; 81pp; English.	
XX	The invention relates to a nucleic acid present in other than its natural	
XX	environment and encoding an Stichodactyliden chromoprotein or its	
CC	fluorescent mutant, where the fluorescent protein has an emission maximum	
CC	ranging from 580-660 nm. The polynucleotides and encoded proteins are	
CC	useful in applications employing a chromo or fluorescent nucleic acid or	
CC	protein. Recombinant vectors comprising the nucleic acid is useful for	
CC	producing an Anthozoan chromo and/or fluorescent protein. The	
CC	chromoproteins, and their fluorescent mutants are useful as colouring	
CC	agents capable of imparting colour or pigment to a particular composition	
CC	of matter. The chromoproteins can be incorporated into a variety of	
CC	different compositions including food compositions, pharmaceuticals,	
CC	cosmetics, living organisms, e.g. animals and plants, and as labels in	
CC	analyte detection assays, e.g. assays for biological analytes of interest	
CC	(see ABL4167 for a detailed description of the various uses of the	
CC	chromoproteins). The present sequence represents the H. crista	
CC	alternative fluorescent protein mutant FP10-crl encoding cDNA	
XX	Sequence 687 BP; 145 A; 230 C; 219 G; 93 T; 0 U; 0 Other;	
Qy	Query Match	59.1%; Score 417.8; DB 6; Length 687;
Db	Best Local Similarity	79.1%; Pred. No. 1.1e-60;
Db	Matches	523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;
Qy	12	CTTGCTGACCGAGACCATGCGCTTCAGGAGCACCATCGAGGGACCGTGAACGGCCACTA 71
Db	12	CTTGCTGAAGGAGCATGCGCATCAAGATGTACATGGAGGACCGTGAACGGCCACTA 71
Qy	72	CTTCAAGTCACCGGCAAGGGCAGGGCAACCCCTCTGAGGGACCCAGAGATGAAGAT 131
Db	72	CTTCAAGTCGAGGGCGAGGGCGACGCGCAACCCCTTCGCCGGCACCCAGAGCATGCGGAT 131
Qy	132	CGAGGTGATCGAGGGCGGGCCCTTCGCGCTTCGCCTCCACATCTGTGCCACCTCTCGAT 191
Db	132	CCAGTGACCGAGGGCGGCCCTTCGCGCTTCGCCTTCGATCTCGATCTCTGCGA 191
Qy	192	GTACGGCTCCAAGGCCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAGAGTC 251
Db	192	GTACGGCAGCAGGACCTTCGTGACCAACACCGCGGAGATCCCGACTCTTCTCAAGCAGAG 251
Qy	252	CCTCCCGAGGGCTTCACCTGGGAGCGCACACACCTACGAGACCGCGGCTTCCTGAC 311
Db	252	CTTCCCGAGGGCTTCACCTGGGAGAGAACCAACCTACGAGACCGCGGCTTCCTGAC 311
Qy	312	CGGCCACGAGACACCTTCCTTGGACGGCGACTGCTGGTGTACAAGGTGAAGATCCTGGG 371
Db	312	CGCCCCACGAGACACCAAGCTTGGAGGCAACTGCTGATCTACAAGGTGAAGGTGCTGGG 371
Qy	372	CAACAACTTCCCGCGACGGCCCGGTGATGTCAGAAACAAGCCGGCCGCTGGGAGCCCTC 431
Db	372	CACCAACTTCCCGCGACGGCCCGGTGATGAAGAAACAAGCGGGCGGCTGGGAGCCCGAG 431
Qy	432	CACCGAGATCGTGTAAGGTGAGACGGCTGTGTCGGCGGCGCAGTCCCTGATGGCCCTTGA 491
Db	432	CACCGAGGTGTTGTAACCGGAGACGGCTGTGTCGGCGGGAACGATGATGGCCCTTGA 491
Qy	492	GTGCCCCCGGGGTCCGACCTGTGACCTGCGACCTGTGACACCACTTACCGCTCCCAAGAGCC 551

Db 492 G---GTGGGCGACCGGGCTGATCTGCCACACTACACAGGTACCGGAGCAAGAAGGC 548
 QY 552 CGCTCCGCGCTGAAGATCCCGGCTTCCACTTCGAGGACACCGCATCGAGATCTCTGGA 611
 Db 549 CGTGGGGCCCTGACCATGCCCGCTTCCACTTCACGACATCGGCTGCAGATGCTGG 608
 QY 612 GGAGGTGAGAGGCGCAAGTCTCAAGACAGTACGAGCGCGCTGGCGCTACTGCGA 671
 Db 609 GAA---GGAGAAGACAGTACTTCGAGCTGTACGAGCGCGTGGCCCGGTACAGCGA 665
 QY 672 C 672
 Db 666 C 666

RESULT 14

ABL41172
 ID ABL41172 standard; cDNA; 587 BP.
 XX
 AC ABL41172;
 DT 12-AUG-2002 (first entry)
 XX
 DE H. crispa fluorescent protein mutant Fp10-cr1 encoding cDNA.

XX Stichodactylidae; chromoprotein; fluorescent; anthozoan; food; FRET;
 KW colouring agent; pharmaceutical; cosmetic; immunoassay; biosensor; gene;
 KW fluorescence resonance energy transfer; gene expression; mutant; ss.
 XX

Heteractis crispa.

XX
 FH Key Location/Qualifiers
 FT CDS 1..687
 FT /*tag= a
 FT /product= "mutant fluorescent protein"

WO200230965-A2.

18-APR-2002.

12-OCT-2001; 2001WO-US032080.

12-OCT-2000; 2000US-0240018P.

16-JUL-2001; 2001US-0306131P.

(CLON-) CLONTECH LAB INC.

Lukyanov SA, Fradkov AF, Lukyanov KA, Gurskaya NG;

WPI; 2002-444170/47.

P-PSDB; ABB07994.

XX Novel nucleic acid encoding Stichodactylidae chromoprotein and its
 FT fluorescent mutant useful as coloring agent, labels in analyte detection
 FT assays, markers in recombinant DNA applications and filters in
 FT sunscreens.

Claim 5; Fig 10; 81pp; English.

XX The invention relates to a nucleic acid present in other than its natural
 CC environment and encoding an Stichodactylidae chromoprotein or its
 CC fluorescent mutant, where the fluorescent protein has an emission maximum
 CC ranging from 580-660 nm. The polynucleotides and encoded proteins are
 CC useful in applications employing a chromo or fluorescent nucleic acid or
 CC protein. Recombinant vectors comprising the nucleic acid is useful for
 CC producing an Anthozoan chromo and/or fluorescent protein. The
 CC chromoproteins, and their fluorescent mutants are useful as colouring
 CC agents capable of imparting colour or pigment to a particular composition
 CC of matter. The chromoproteins can be incorporated into a variety of
 CC different compositions including food compositions, pharmaceuticals,
 CC cosmetics, living organisms, e.g. animals and plants, and as labels in
 CC analyte detection assays, e.g. assays for biological analyses of interest
 CC (see ABL41167 for a detailed description of the various uses of the

CC chromoproteins). The present sequence represents the H. crispa
 CC fluorescent protein mutant Fp10-cr1 encoding cDNA

XX Sequence 687 BP; 150 A; 232 C; 212 G; 93 T; 0 U; 0 Other;

XX Query Match 59.1%; Score 417.8; DB 6; Length 687;

XX Best Local Similarity 79.1%; Pred. No. 1.1e-60;

XX Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;

QY 12 CTGTGTGACCGAGACCATGCTTCCAGCACCATCGAGGCGACCGTGAACGGCCACTA 71

Db 12 CTGTGTGAGAGAGATGCGCATCAAGATGATGAGGGGCGACCGTGAACGGCCACTA 71

QY 72 CTTCAAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGGCGACCGAGATGAAGAT 131

Db 72 CTTCAAGTGCAGGGCGGAGGCGGCGCAACCCCTCGAGGGGCGACCGAGATGAAGAT 131

QY 132 CGAGGTGATCGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 191

Db 132 CGAGGTGATCGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 191

QY 192 GTACGGCTCCAAAGCGCTTTCATCAAGTAGCTGTCCGGCATCCCGACTACTTCAAGCAGTC 251

Db 192 GTACGGCAGCAGGAGGCTTTCATCAAGTAGCTGTCCGGCATCCCGACTACTTCAAGCAGAG 251

QY 252 CTCCCGGAGGGCTTCACTGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 311

Db 252 CTCCCGGAGGGCTTCACTGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 311

QY 312 CGCCACCGAGGACACCTCCCTGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 371

Db 312 CGCCACCGAGGACACCTCCCTGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 371

QY 372 CAACAACTTCCCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 431

Db 372 CAACAACTTCCCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 431

QY 432 CACCGAGATCGTGTACGAGGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 491

Db 432 CACCGAGGTGTGTACCGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 491

QY 492 GTGCGCGCGGCTGCCACCTGACCTGCCACCTGCCACCTGCCACCTGCCACCTGCCAC 551

Db 492 G---GTGGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 548

QY 552 CGCCTCCGCGCTGAGATGCCGCTTCCACTTCGAGGAGCGGCGGCGGCGGCGGCGGCGG 611

Db 549 CGTGGCGCGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 608

QY 612 GGAGGTGAGAGGCGGCAAGTGTCTACAAGCAGTACGAGGCGGCGGCGGCGGCGGCGG 671

Db 609 GAA---GAAGAAGGACGAGTACTTCGAGCTGTACGAGGCGGCGGCGGCGGCGGCGG 665

672 C 672

666 C 666

RESULT 15

AAI69492
 ID AAI69492 standard; cDNA; 955 BP.

XX

AC AAI69492;

DT 12-MAR-2002 (first entry)

XX

DE A. sulcata asCP562 cDNA.

XX

KW Marker protein; green fluorescent protein; GFP; transgenic plant;

XX orange fluorescent protein; ss.

XX Anemonia sulcata.

OS

XX

Key	Location/Qualifiers
FD	26..472
FT	/*tag= a
FT	/product= "asCP562"
XX	
XX	DE10026520-A1.
XX	
XX	13-DEC-2001.
XX	
XX	27-MAY-2000; 2000DE-01026520.
XX	
XX	27-MAY-2000; 2000DE-01026520.
XX	
XX	(WIED/) WIEDENMANN J.
XX	
XX	Wiedenmann J, Daeschner K, Binder S;
XX	
XX	WPI; 2002-091071/13.
XX	
XX	New fluorescent or colored proteins from Anemonia sulcata, useful as markers in molecular biology and genetic engineering.
XX	
XX	Disclosure; Page 3; 6pp; German.
XX	
XX	This invention describes novel proteins from Anemonia sulcata characterised by green or orange fluorescence or a natural red colouration. The products of the invention, or their mutants, are used as markers in molecular biology and genetic engineering, or generally wherever their fluorescent/coloured properties are useful. The green fluorescent protein from A. sulcata is better suited to use in transgenic plants than known fluorescent proteins, particularly it has stronger fluorescence than the green fluorescent protein from Aequorea victoria. This sequence encodes an A. sulcata derived fluorescent protein which is described in the invention
XX	
XX	Sequence 955 BP; 277 A; 194 C; 215 G; 269 T; 0 U; 0 Other;
XX	
XX	Query Match 59.1%; Score 417.8; DB 6; Length 955;
XX	Best Local Similarity 75.9%; Pred. No. 1.1e-60;
XX	Matches 529; Conservative 0; Mismatches 167; Indels 1; Gaps 1;
QY	7 GCCTCCCTGCTGACCGAGACCATGCCTTCAGGACCAACCATCGAGGGCACCTGTGAACGGC 66
Db	29 GCTTCTCTTTTAAAGAACTATGCCTTTTAAAGAGACCATTAAGGGACGGTTAATGGC 88
QY	67 CACTACTTTCAAGTGACCGGAGGGGAGGGGACCCCTCGAGGGCACCGAGGATG 126
Db	89 CACTACTTTCAAGTGTACAGGAAAGAGAGGGCAACCCATTTGAGGGTACGAGGAATG 148
QY	127 AAGATCGAGGTGATCGAGGGGGCCCTCGCCCTTCGCTTCACATCTCTGTCCACCTCC 186
Db	149 AAGATAGAGGTGATCGAGAGGAGTCCATTGCCATTTGCCATTTGTCAACGAGT 208
QY	187 TCGATGTACGGCTCCAGGCCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db	209 TGTATGTACGGTGTAGTAAAGACCTTCATCAAGTACGTGTTCAGGAAATCCAGACTTCAAG 268
QY	247 CAGTCCCTCCCGAGGGCTTCACCTGGAGGGCACCAACCACTACAGAGACGGCGCTTC 306
Db	269 CAGTCTTTCCTGAGGTTTTACTTTGGAAAGAACCAACCTACGAGGATGAGGCTTT 328
QY	307 CTGACCGCCCAACAGGACACTTCCTGGAGCGCGACTGCTGTGTGTACAGGTGAAGATC 366
Db	329 CTTACAGCTCATCAGGACACAGGCTAGATGGAGATTGCCCTCGTTTACAGGTCAAGATT 388
QY	367 CTGGGCAACAACTTCCCGCGCGACGG - CCGCGTGCATGCAGAACAAAGCGCGCGTGGGA 425
Db	389 CTTGTGTAATAATTTCTGTGATGGCCCCCGTGTGAGAAACAGGACGAAGATGGGA 448
QY	426 GCCCTCCACCGAGATCGTGTACGAGGTGGACGGCGTCTGCGGGCGAGTCCCTGTATGGC 485
Db	449 GCCAGCACCGAGATACTTTATGAAGTGAAGGTGTCTCGTGGACAGTCTTTGATGGC 508

Blank Sheet as PTD

Db 1 AlaSerLeuLeuThrGluThrMetProPheAArgThrThrIleGluGlyThrValAsnGly 20
 QY 67 CACTACTCAAGTGCACCGGAGGCGGAGGCAACCCCTCGAGGGGACCCAGAGATG 126
 Db 21 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGlnGluMet 40
 QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTCGCTTCCGCTTCCATCTGTCACCTCC 186
 Db 41 LysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
 QY 187 TGCATGTACGGTCCCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
 Db 61 CysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
 QY 247 CAGTCCCTCCCGAGGCTTCACCTGGAGCGGCGGACACACCTACGAGGACGGCGCTC 306
 Db 81 GlnSerLeuProGluGlyPheThrTrpGluA-gThrThrTyrGluAspGlyGlyPhe 100
 QY 307 CTGACGCGCCACGAGACACCTCCCTGGAGCGGCGACTGCTGTGTACAGGTGAAGATC 366
 Db 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValIle 120
 QY 367 CTGGGCAACAACTTCCCGCGGAGCGGCGGCTGATCGAGAACAGGCGCGCGCTGGAG 426
 Db 121 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 140
 QY 427 CCTCCACGAGATCGTACGAGGTGGAGCGGCTGCTGGCGGCGGCTCCCTGATGGCC 486
 Db 141 ProSerThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
 QY 487 CTGAGTGGCCCGCGGCTGCCACTGACCTGCACCTGCACACCACTACGCTCCAG 546
 Db 161 LeuGluCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 180
 QY 547 AAGCCGCTCCCGCTGAGATGCGCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
 Db 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 200
 QY 607 CTGAGGAGGTGAGAGGCAAGTCTCAAGCAGTACGAGGCGGCGGCTGGCGCGCTAC 666
 Db 201 LeuGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 220
 QY 667 TGGAGCGCCCGCTCCCAAGTGGGCGCAAC 699
 Db 221 CysAspAlaAlaProSerLysLeuGlyHisAsn 231

RESULT 3

ABU09924

ID ABU09924 standard; protein; 468 AA.

XX AC ABU09924;

XX XX 09-JUL-2003 (first entry)

XX XX 17-APR-2003

XX XX Chromoprotein AsRed-35-SNA-tandem.

XX XX Chromo/fluorescent domain; labeled fusion protein;

XX KW site-specific gene modification; chromoprotein; colouring agent;

XX KW food composition; pharmaceutical; cosmetic; AsRed-35-SNA-tandem.

XX XX Authorza.

OS

XX WO2003031590-A2.

XX XX 17-APR-2003.

XX XX 10-OCT-2002; 2002WO-US032560.

XX XX 12-OCT-2001; 2001US-00976673.

XX XX 11-FEB-2002; 2002US-0356225P.

XX XX 22-MAY-2002; 2002US-0383336P.

XX XX

(CLON-) CLONTECH LAB INC.

XX Lubyakov SA;

XX WPI; 2003-381709/36.

XX N-PSDB; ACA61027.

XX New nucleic acid encoding polypeptide products having at least two linked
 chromo/fluorescent domains, useful for generating transgenic plants or
 animals or site-specific gene modifications in cell lines.

XX Disclosure; Fig 4; 68pp; English.

XX The invention describes a nucleic acid encoding a polypeptide product
 comprising a first and a second chromo/fluorescent domain, optionally
 joined by a linking domain. The first and second chromo/fluorescent
 domains associate with each other under intracellular conditions so that
 the encoded polypeptide assumes a tertiary structure. The nucleic acid
 and the protein are useful in producing labeled fusion proteins that have
 a precise and predictable signal to fusion partner ratio. The nucleic
 acid may also be used in generating transgenic, non-human plants or
 animals or site-specific gene modifications in cell lines. The
 chromoproteins may be used as colouring agents, as a food composition, in
 pharmaceuticals or cosmetics, as labels in analyte detection assays or as
 selectable markers in recombinant DNA applications. This is the amino
 acid sequence of chromo/fluorescent domain fusion protein AsRed-35-SNA-
 tandem

XX Sequence 468 AA;

Alignment Scores:

Pred. No.: 1.09e-89 Length: 468
 Score: 1256.00 Matches: 231
 Percent Similarity: 99.14% Conservative: 0
 Best Local Similarity: 99.14% Mismatches: 2
 Query Match: 96.62% Indels: 0
 DB: Gaps: 0

US-10-081-864-21 (1-707) x ABU09924 (1-468)

QY 1 GATCCGCTCCCTCTGCTACGAGACCATGCTTACGACACCATCGAGGGACCGTG 60
 Db 236 GlyMetAlaSerLeuLeuThrGluThrMetProPheArgThrThrIleGluGlyThrVal 255
 QY 61 AACGCCCACTACTTCAAGTGCACCGGCAAGGGCGGAGGCAACCCCTCGAGGGACCCAG 120
 Db 256 AsnGlyHisTyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGln 275
 QY 121 GAGATGAAGATCGAGGTGATCGAGGCGGCGGCTTCCGCTTCCCATCTCTGTCC 180
 Db 276 GluMetLysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSer 295
 QY 181 ACCTCTGCATGATCGGTCCAAAGCCCTTCATCAAGTACGTGTCGGGATCCCCGACTAC 240
 Db 296 ThrSerCysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyr 315
 QY 241 TTCAAGCAGTCCCTCCCGAGGGCTTCACTGGAGCGCACCCACCTACGAGACGGC 300
 Db 316 PheLysGlnSerLeuProGluGlyPheThrTrpGluA-gThrThrTyrGluAspGly 335
 QY 301 GGCTTCTGACCGCCCAACGAGACACCTCCCTGAGCGGCGACTGCTGCTGTACAGGTG 360
 Db 336 GlyPheLeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysVal 355
 QY 361 AAGATCTGGGCAACAACCTTCCCGCGGCGGCTTCACTGGAGCGCACCCACCAAGCGCGCGC 420
 Db 356 LysIleLeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArg 375
 QY 421 TGGAGAGCCCTCCACGAGATCGTGTACGAGGTGAGCGGCTGCTCGCGGCGGCGCTCG 480
 Db 376 TrpGluProSerThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerSer 395
 QY 481 ATGGCCCTGGAGTGCGCGCGGCTGCGCACCTGACCTGCCACCTGCACACCACTACCGC 540

PD 15-JUN-2000.
XX
PF 10-DEC-1999; 99WO-US029300.
XX
PR 11-DEC-1998; 98US-00210330.
XX
PR 09-DEC-1999; 99US-00210330.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukanoy SA, Fradkov AF, Labas YA, Matz MV, Jiang X, Duong T;
PI Zhao X;
XX WPI: 2000-423374/36.
XX N-PSDB; AA050884.
XX
PT Novel fluorescent proteins from non-bioluminescent Anemonia sulcata,
PT useful for fluorescent labeling and as markers.
XX
XX Claim 9; Page 66-67; 71pp; English.
XX
XX The present sequence is the mutant fluorescent protein, Mut1, from non-
CC bioluminescent Anemonia sulcata. The fluorescent protein has applications
CC in fluorescent labelling, as fluorescent markers for gene expression and
CC protein localisation studies, and in fluorescence resonance energy
CC transfer (FRET) reactions. The coding sequence may be used as a source or
CC primers and probes for identifying related proteins
XX
XX Sequence 231 AA;
SQ

Alignment Scores:
Pred. No.: 2,07e-87 Length: 231
Score: 1226.00 Matches: 223
Percent Similarity: 98.27% Conservative: 4
Best Local Similarity: 96.54% Mismatches: 4
Query Match: 94.31% Indels: 0
DB: Gaps: 0

US-10-081-864-21 (1-707) x AAY99734 (1-231)

QY 7 GCTCCCTGCTGACCGAGACCATGCTTTCAGGACACCATCGAGGACACCGTGAACGCG 66
Db 1 AlaserPheLeuLysThrMetProPheLysThrThrilleGluGlyThrValasngly 20
QY 67 CACTACTTCAAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGACCCAGGAGATG 126
Db 21 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 40
QY 127 AGATCGAGGTGATCGAGGCGGCGCCCTGCGCTTCCGCTTCCACATCCTGTCCACCTCC 186
Db 41 LysileGluValileGluGlyProLeuProPheAlaPheHisileLeuSerThrSer 60
QY 187 TGATGTACGGCTCCCAAGGCTTCATCAAGTAGCTGCTCGGATCCCGACTACTTCAAG 246
Db 61 CysMetTyGlySerLysAlaPheileLysTyValSerGlyleProaspTyPheLys 80
QY 247 CAGTCCCTCCCGAGGGCTTACCTGGAGCGGCACCCACCATCGAGGACGGCGCTTC 306
Db 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACCGAGACACCTCCCTGGACGGGACTGCTGTGTACAGGTGAAGATC 366
Db 101 LeuthAlahisGlnaspThrSerLeuAspGlyAspCysLeuValTyLysVallylle 120
QY 367 CTGGGCAACAACCTCCCGCGAGCGCCCGTGTATGCAGACAAGCGCGCGCTGGGAG 426
Db 121 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 140
QY 427 CCTTCCACCGAGATCGTGTACAGGTGAGCGGCTGCTGCGGCGGCGCTGTGGCC 486
Db 141 ProSerThrGluileValTyGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
QY 487 CTGGAGTGCCTCCCGGCTGCCACCTGACCTGCGACCTGCACACCACTACCGCTCCAG 546

Db 161 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyArgSerLys 180
QY 547 AAGCCGCTCCGCTCCGCTGAAAGATGCCGCTTCCACTTCGAGGACACCGCATCGAGATC 606
Db 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgileGluile 200
QY 607 CTGAGGAGGTGAGAGGCAAGTGTCTCAAGCAGTACGAGGCGCGCTGGCGCGCTAC 666
Db 201 MetGluGluValGluLysGlyLysCysTyLysGlnTyGlnAlaAlaValGlyArgTy 220
QY 667 TGGCAGCGCCGCTCCCAAGCTGGGCGCAAC 699
Db 221 CysAspAlaAlaProSerLysLeuGlyHisAsn 231
RESULT 6
AAE00376
ID AAE00376 standard; protein; 232 AA.
XX
XX AAE00376;
XX
XX 19-JUN-2001 (first entry)
DT
XX
XX Anemonia sulcata chromo/fluorescent protein, asFP600 (NFP-7).
DE
XX Anthozoa; Chromoprotein; fluorescent protein; asFP600; NFP-7; sunscreen;
KW analyte detection assay; selectable marker; recombinant DNA application;
KW biosensor; pH indicator; invivo marker; selective filter; sea anemone.
XX
XX Anemonia sulcata.
XX
XX WO200127150-A2.
XX
XX 19-APR-2001.
XX
XX 13-OCT-2000; 2000WO-US028477.
XX
XX 14-OCT-1999; 99US-00418529.
PR 15-OCT-1999; 99US-00418517.
PR 15-OCT-1999; 99US-00418522.
PR 19-NOV-1999; 99US-0044338.
PR 19-NOV-1999; 99US-0044341.
PR 09-DEC-1999; 99US-00457556.
PR 09-DEC-1999; 99US-00457898.
PR 09-DEC-1999; 99US-00458144.
PR 09-DEC-1999; 99US-00458477.
PR 10-DEC-1999; 99WO-US029405.
PR 14-JUN-2000; 2000US-0211607P.
PR 14-JUN-2000; 2000US-0211609P.
PR 14-JUN-2000; 2000US-0211626P.
PR 14-JUN-2000; 2000US-0211627P.
PR 14-JUN-2000; 2000US-0211687P.
PR 14-JUN-2000; 2000US-0211766P.
PR 14-JUN-2000; 2000US-0211880P.
PR 14-JUN-2000; 2000US-0211888P.
PR 14-JUN-2000; 2000US-0212070P.
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukanoy SA, Fradkov AF, Labas YA, Matz MV, Tersikh A;
XX
XX WPI; 2001-266409/27.
XX N-PSDB; AAD03616.
XX
XX An Anthozoa chromo- or fluorescent protein (PI) present in an environment
PT other than its natural environment, useful as a label in analyte
PT detection assays and as a selectable marker in recombinant DNA
PT applications.
XX
XX Claim 8; Fig 7; 69pp; English.
XX
XX The present sequence is an Anemonia sulcata chromo/fluorescent protein,
CC asFP600 (NFP-7). NFP-7 is present in an environment other than its
CC natural environment and has an absorbance maximum ranging from 370nm to

CC 770nm and more usually from 560nm to 580nm and emission maximum ranging
 CC from 395nm to 795nm and more usually from 585 to 605nm. The
 CC chromoproteins or fluorescent proteins are useful as labels in analyte
 CC detection assays, as selectable markers in recombinant DNA applications,
 CC as biosensors in prokaryotic and eukaryotic cells e.g. as pH indicator
 CC and as in vivo markers in animals. They are also useful in sunscreens and
 CC as selective filters. Note: The specification also refers to a mutant of
 CC this sequence, Mut1 with substitutions T70A and A148S. However, A at
 CC position 148 do not match with this wild type sequence
 XX
 SQ Sequence 232 AA;

Alignment Scores:
 Pred. No.: 3,56e-87 Length: 232
 Score: 1223.00 Matches: 222
 Percent Similarity: 98.27% Conservative: 5
 Best Local Similarity: 96.10% Mismatches: 4
 Query Match: 94.08% Indels: 0
 DB: 4 Gaps: 0

US-10-081-864-21 (1-707) x AAE00385 (1-232)

QY 7 GCCTCCCTGTCAGCCGAGACATGCTCCCTCAGGACACCATCAGGCGCACCGTGAACGGC 66
 Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrileGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTCACCGCAGGCGGAGGCAACCCCTCAGGCGCACCGTGAACGGC 126
 Db 22 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGTCAGGTGATCGAGCGCGCGCCCTCCCTCCCTCCACATCCTGTCACCTCC 186
 Db 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
 QY 187 TGCATGTACGGCTCCAGGCGCTTCATCAGTGTCTCGGCTCCGCGCATCCGCTACTTCAG 246
 Db 62 CysMetTyrGlySerLysAlaPheileLysTyrValSerGlyProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGCGCTTCACCTGGGAGCGCACCCACCATCAGGCGCGCGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACAGGACACCTCCCTGGAGCGGCGCTGCTGTGTACAGGTGATGATC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysile 121
 QY 367 CTGGCAACAACTCCCGCGGCGCGCTGATGAGAACAGCGCGCGCTGGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTCACCAGATCGTGTACAGGTGAGCGCGCTGCTGGCGCGCGCTGATGATGCGC 486
 Db 142 ProSerThrGluileValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGGAGTCCCGCGCGCTGCGCACCTGACCTGCGCACCTACCGCTCCCGAG 546
 Db 162 LeuLysCysProGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCCCGCTCCCGCTGAGATGCGCGCTTCACCTCAGGACACCGCATCGATGATC 606
 Db 182 LysProAlaAlaLeuLysMetProGlyPheHisPheGluAspHisArgileGluile 201
 QY 607 CTGGAGGTGAGAGGCGAGTCTACAGCAGTACAGCGCGCGCGCTGCGCGCTAC 666
 Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
 QY 667 TGGAGCGCGCGCGCTCCAGCTGGGCGCGCAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 7

AAE00385

ID AAE00385 standard; protein; 232 AA.

XX AAE00385;
 AC 19-JUN-2001 (first entry)
 DT Anemonia sulcata asPP600 (NFP-7) mutant (A184S).
 XX Anthozoa; Chromoprotein; fluorescent protein; asPP600; NFP-7; sunsreen;
 KW analyte detection assay; selectable marker; recombinant DNA application;
 KW biosensor; pH indicator; invivo marker; selective filter; sea anemone;
 KW mutant; mutein.
 OS Anemonia sulcata.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 184
 FT /note= "Wild type Ala substituted with Ser"
 XX
 PN WO200127150-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 13-OCT-2000; 2000WO-US028477.
 XX
 PR 14-OCT-1999; 99US-00418529.
 PR 15-OCT-1999; 99US-00418917.
 PR 15-OCT-1999; 99US-00418922.
 PR 19-NOV-1999; 99US-00444338.
 PR 19-NOV-1999; 99US-00444341.
 PR 09-DEC-1999; 99US-00457556.
 PR 09-DEC-1999; 99US-00457898.
 PR 09-DEC-1999; 99US-00458144.
 PR 09-DEC-1999; 99US-00458477.
 PR 10-DEC-1999; 99WO-US029405.
 PR 14-JUN-2000; 2000US-0211607P.
 PR 14-JUN-2000; 2000US-0211609P.
 PR 14-JUN-2000; 2000US-0211626P.
 PR 14-JUN-2000; 2000US-0211627P.
 PR 14-JUN-2000; 2000US-0211687P.
 PR 14-JUN-2000; 2000US-0211766P.
 PR 14-JUN-2000; 2000US-0211880P.
 PR 14-JUN-2000; 2000US-0211888P.
 PR 14-JUN-2000; 2000US-0212070P.
 XX
 PA (CLON-) CLONTECH LAB INC.
 XX
 PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Tersikh A;
 XX WPI; 2001-265409/27.
 DR
 PT An Anthozoa chromo- or fluorescent protein (P1) present in an environment
 PT other than its natural environment, useful as a label in analyte
 PT detection assays and as a selectable marker in recombinant DNA
 PT applications.
 XX
 PS Disclosure; Page: 69pp; English.
 XX
 CC The present sequence is an Anemonia sulcata chromo/fluorescent protein,
 CC asPP600 (NFP-7) mutant (A184S). NFP-7 is present in an environment other
 CC than its natural environment and has an absorbance maximum ranging from
 CC 370nm to 770nm and more usually from 560nm to 580nm and emission maximum
 CC ranging from 395nm to 795nm and more usually from 585 to 605nm. The
 CC chromoproteins or fluorescent proteins are useful as labels in analyte
 CC detection assays, as selectable markers in recombinant DNA applications,
 CC as biosensors in prokaryotic and eukaryotic cells e.g. as pH indicator
 CC and as in vivo markers in animals. They are also useful in sunscreens and
 CC as selective filters. Note: The present sequence is not shown in the
 CC specification, but derived from asPP600 (NFP-7) sequence (AAE00376) shown
 CC in figure 7
 XX
 SQ Sequence 232 AA;

XX 01-MAR-2002; 2002WO-GB000928.
 XX 02-MAR-2001; 2001US-0273227P.
 PR 21-MAR-2001; 2001AU-00003874.
 PR 15-OCT-2001; 2001US-0329816P.
 XX (NUFA-) NUFARM LTD.
 PA (UYOU) UNIV QUEENSLAND.
 PA (JONE/) JONES E L.
 XX Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
 PI Hoegh-Guldberg IO, Prescott M;
 PI WPI; 2002-740765/80.
 XX Novel color-facilitating molecule for producing a biomatrix, has a
 PT polypeptide which alone/along with molecules imparts altered visual
 PT characteristics to cells in the absence of excitation by extraneous non-
 PT white light.
 XX Example 20; Page 503-504; 510pp; English.
 XX The invention relates to an isolated colour-facilitating molecule (CFM)
 CC comprising a polypeptide which, in a cell, alone or together with one or
 CC more other molecules imparts an altered visual characteristic to the cell
 CC when visualised by a human eye in the absence of excitation by extraneous
 CC non-white light or particle emission. CFMs are useful for producing a
 CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
 CC red coloured fleece. They are useful for producing coloured plant
 CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
 CC uses include transducing or intensifying an image, providing additional
 CC light for growing phototropic organisms e.g. algae and/or corals, for
 CC coating materials that experience UV damage e.g. plastics and car
 CC upholstery. CFMs are useful in the flower industry, in the development of
 CC new varieties of flowering plants. Other contemplated uses include,
 CC expression markers, general reporter molecules, photon traps, UV sinks or
 CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
 CC fungal species, and in fruits and vegetables to enhance their
 CC marketability. CFMs embedded in a gel matrix improve image quality in
 CC situations of distorted light spectra (biomatrix). The first all-protein
 CC chromophore to be isolated was Green Fluorescent protein (GFP). The
 CC sequences given in records ABP69924-ABP70048 represent CFM related amino
 CC acid sequences
 XX Sequence 232 AA;
 SQ

Alignment Scores:
 Pred. No.: 7, 31e-87 Length: 232
 Score: 1219.00 Matches: 221
 Percent Similarity: 97.84% Conservatives: 5
 Best Local Similarity: 95.67% Mismatches: 5
 Query Match: 93.77% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-21 (1-707) x ABP70043 (1-232)

QY 7 GCTCTCTGTCAGCAGACCATGCTTCAGGACCAACCATGAGGCGACCGTGAACGCG 66
 Db 2 AlaserPheLeuLysLysThrMetProPheLysThrIleGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTGCACCGCAGGCGGAGGCGCAACCCCTCGAGGGCACCCAGGAGATG 126
 Db 22 HistyPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCGAGGTGATCGAGGCGCGCCCTGCTTCACATCTGCTCCACCTCC 186
 Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
 QY 187 TGCATGTACGCTCCAGGCGCTTCATCAAGTAGCTGTCGGCATCCCGCACTACTCAAG 246
 Db 62 CysMetCysGlySerLysThrPheIleLysThrValSerGlyLeuProAspTyrPheLys 81

QY 247 CAGTCCCTCCCGAGGGCTTCACTGGAGCGCACCCACCTACGAGACCGCGGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACACGAGACACCTCCCTGGAGCGGAGCTGCTGTGTACAAAGGTGAAGATC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
 QY 367 CTGGCAACAACATCCCGCCGCGACGCGCGCTGATCAGAACAAAGCCGCGCTGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTTCACCGAGATCGTGTACGAGGTGGAGCGGCTGCTGGCGCCAGTCCCTGATGGCC 486
 Db 142 ProAlaThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGGAGTGCCTCCCGCGCTGCGCACCTGACCTGCGCACACCTACCTCCCTCAAG 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCCGCTCCCGCTGAAAGATGCCGCGCTTCACCTTCGAGGACCAACCGCATCGATC 606
 Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
 QY 607 CTGGAGGAGTGGAGAGGCAAGTGTCTACAAGCAGTACGAGGCGCGCTGGCGCGCTAC 666
 Db 202 MetGluGluValGlnLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
 QY 667 TGGAGCCCGCCCTCCCAAGCTGGGCGCAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 11
 AAG79762
 ID AAG79762 standard; protein; 232 AA.
 AC AAG79762;
 XX 01-APR-2003 (first entry)
 DT Wild type kindling fluorescent protein, AsFP595.
 XX Kindling fluorescent protein; kindling stimulus; movement; labeling;
 DE fluorescence resonance energy transfer; FRET;
 XX bioluminescence resonance energy transfer; BRET; biosensor;
 KW automated screening.
 KW Anthozoa sulcata.
 XX OS
 XX WO200296924-A1.
 XX PD
 XX 05-DEC-2002.
 XX PF
 XX 24-MAY-2002; 2002WO-US016379.
 XX PR
 XX 25-MAY-2001; 2001US-0293752P.
 XX PR
 XX 11-OCT-2001; 2001US-0329176P.
 XX PA
 XX (CLON-) CLONTECH LAB INC.
 XX PI
 XX Lukyanov SA, Chudakov D, Lukyanov X;
 XX WPI; 2003-156789/15.
 XX DR
 XX N-PSDB; ABA00799.
 XX PT
 XX Novel nucleic acid that is present in other than its natural environment
 PT and that encodes kindling fluorescent protein, is useful in labeling
 PT protocols, e.g. labeling proteins, organelles, cells and organisms.
 XX Example; Fig 1; 96pp; English.
 XX The sequences given in AAG79762-69 show wild type and mutant kindling
 CC fluorescent proteins. The proteins go from a first substantially non-

CC fluorescent or non-fluorescent state to a second fluorescent state upon
 CC exposure to a kindling stimulus. The kindling proteins are useful for
 CC detecting an entity such as a protein, organelle or cell in a composition
 CC such as a cell or a multicellular composition (preferably a multicellular
 CC organism), by providing the entity as an entity labeled with the kindling
 CC protein, kindling the kindling fluorescent protein label with a kindling
 CC stimulus to produce a kindling fluorescent protein label, and
 CC exciting the kindling fluorescent protein label with light and
 CC detecting any fluorescence from it to detect the entity. The method
 CC monitors the movement of the entity. The fluorescent proteins and the
 CC cDNA encoding them are useful in labeling protocols, e.g., labeling
 CC proteins, organelles, cells and organisms as biological labels or
 CC markers, in protein labeling or tagging applications. The fluorescent
 CC kindling proteins are useful as detectable labels, as labels in analyte
 CC detection assays, in fluorescence resonance energy transfer (FRET)
 CC applications, as biosensors in prokaryotic and eukaryotic cells, in
 CC applications involving the automated screening of arrays of cells
 CC expressing fluorescent reporting groups, in high through-put screening
 CC assays, as second messenger detectors, and in fluorescent activated cell
 CC sorting assays

XX SQ Sequence 232 AA;

Alignment Scores:
 Pred. No.: 7, 31e-87 Length: 232
 Score: 1219.00 Matches: 221
 Percent Similarity: 97.84% Conservatives: 5
 Best Local Similarity: 95.67% Mismatches: 5
 Query Match: 93.77% Indels: 0
 DB: Gaps: 0

US-10-081-864-21 (1-707) x AAG79762 (1-232)

QY 7 GCTCCCTGTCACGAGACCATGCGCTTCAGGACCAACCATCGAGGCGACCGTGAACGCG 66
 Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValaenGly 21
 QY 67 CACTACTTCAAGTGACCGGACGAGGCGAGGCAACCCCTCGAGGCGACCCAGGAGATG 126
 Db 22 HisTy-PhelysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AGATGAGTGATGACGAGGCGCGCCCTTCGCTTCACATCCGTCGCCACCTCC 186
 Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
 QY 187 TGCATGTACGCTCCAGGCGCTTCATCAAGTACGTGTCGGCATCCCGACTACTTCAAG 246
 Db 62 CysMetTyGlySerLysThrPheIleLysTyValSerGlyIleProAspTyPheLys 81
 QY 247 CAGTCCCTCCCGAGGCGTTCACCTCGGAGCGCACCCACCATCTAGGACGCGCGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTy-GluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACGACACCTCCCTGGAGCGGACTGCTGCTGTACAGGTGAATC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrlsVallysile 121
 QY 367 CTGGGCAACAACTTCCCGCGACGCGCCCGGTGATGACAGAACGCGCGCGTGGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTCCACGAGATGCTGTACAGGTGAGCGGCTGCTGCGCGGCGAGTCCCTGATGCC 486
 Db 142 ProAlaThrGluIleValTyrlsGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGGAGTCCCGCGCGTCCACCTGACCTCCACCTGACACCATCTACGCTCCCAAG 546
 Db 162 LeuLysCysProGlyArgHisLeuThrCysHisLeuHisThrThrTyArgSerLys 181
 QY 547 AGCCCGCTCCGCTGAGATGCGCGCTTCCACTTGGAGACCAACCGCATCGATC 606
 Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201

QY 607 CTGAGGAGGTGGAGAGGCAAGTGTCTAAGAGTACGAGCGCGCGTGGCGGTAC 666
 Db 202 MetGluGluValGluLysGlyLysCysTyrlsGlnTyrlsGluAlaValGlyArgTyrls 221
 QY 657 TCGACGCGCGCGCGCTCCCAAGCTGGGCGCACAAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 12

AAG79763

ID AAG79763 standard; protein; 232 AA.

XX AAG79763;

XX 01-APR-2003 (first entry)

XX Kindling fluorescent protein, AsP595 Al48G.

XX Kindling fluorescent protein; kindling stimulus; movement; labeling;

KW fluorescence resonance energy transfer; FRET;

KW bioluminescence resonance energy transfer; BRET; biosensor;

KW automated screening.

XX Anthozoa sulcata.

XX Key Location/Qualifiers

FT Misc-difference 148

FT /label= Al48G

XX WO200296924-A1.

XX 05-DEC-2002.

XX 24-MAY-2002; 2002WO-US016379.

XX 25-MAY-2001; 2001US-0293752P.

PR 11-OCT-2001; 2001US-0329176P.

XX (CLON-) CLONTECH LAB INC.

PI Lukyanov SA, Chudakov D, Lukyanov K;

DR WPI; 2003-156789/15.

DR N-PSDB; ABA00800.

XX Novel nucleic acid that is present in other than its natural environment and that encodes kindling fluorescent protein, is useful in labeling protocols, e.g. labeling proteins, organelles, cells and organisms.

XX Example; Fig 2; 96pp; English.

The sequences given in AAG79762-69 show wild type and mutant kindling fluorescent proteins. The proteins go from a first substantially non-fluorescent or non-fluorescent state to a second fluorescent state upon exposure to a kindling stimulus. The kindling proteins are useful for detecting an entity such as a protein, organelle or cell in a composition such as a cell or a multicellular composition (preferably a multicellular organism), by providing the entity as an entity labeled with the kindling protein, kindling the kindling fluorescent protein label with a kindling stimulus to produce a kindling fluorescent protein label, and exciting the kindling fluorescent protein label with light and detecting any fluorescence from it to detect the entity. The method monitors the movement of the entity. The fluorescent proteins and the cDNA encoding them are useful in labeling protocols, e.g., labeling proteins, organelles, cells and organisms as biological labels or markers, in protein labeling or tagging applications. The fluorescent kindling proteins are useful as detectable labels, as labels in analyte detection assays, in fluorescence resonance energy transfer (FRET) applications, as biosensors in prokaryotic and eukaryotic cells, in applications involving the automated screening of arrays of cells expressing fluorescent reporting groups, in high through-put screening

CC assays, as second messenger detectors, and in fluorescent activated cell
 CC sorting assays
 XX Sequence 232 AA;

Alignment Scores:
 Pred. No.: 8.75e-87 Length: 232
 Score: 1216.00 Matches: 221
 Percent Similarity: 97.40% Conservatives: 4
 Best Local Similarity: 95.67% Mismatches: 6
 Query Match: 93.69% Indels: 0
 DB: 6 Gaps: 0

US-10-081-864-21 (1-707) x AAG79763 (1-232)

```

QY 7 GCCTCCCTGTCAGCAGACACATCCCTTCAGACACACATCCAGGCGCCGCTGAACGCG 66
DB 2 AlaserPheLeuLysThrMetProPheLysThrThrileGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTCCACCGGCAAGGCGAGGCAACCCCTCGAGGCGCACCCAGGAGATG 126
DB 22 Histy-PheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCGAGGTATCGAGGCGGCGCCCTCGCTTCCCTTCACATCTCGTCCACTCC 186
DB 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCGGCATCCCGCACTACTTCAAG 246
DB 62 CysMetTyrgLysSerLysThrPheLysTyrsValserGlyleProAspTyrgLys 81
QY 247 CAGTCCCTCCCGAGGCTTCACCTGGAGCGGCACACACACTTACGAGACGCGGCTTC 306
DB 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrgLysPheGlyPhe 101
QY 307 CTGACCGGCCACACAGGACACCTCCCTCGACGGGACTCCCTGCTGTACAGGTGAAGATC 366
DB 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrgLysVallylle 121
QY 367 CTGGGCAACACTTCCCGCGGAGCGCCCGTGTATGAGAACAGCGCGCGCTGGAG 426
DB 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
QY 427 CCCTCCACCGAGATCGTGTAGAGTGGAGCGGTGCTGGCGCGCCAGTCCCTGTATGGC 486
DB 142 ProGlyThrGluLysValTyrgLysValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGAGATGCCCGCGGTGCGCACTGACCTGACCTGACACCTGACACCTACCGCTCCAA 546
DB 162 LeuLysCysProGlyGlyA-gHisLeuThrCysHisLeuHisThrThrTyrgSerLys 181
QY 547 AAGCCGCTCCCGCTGAAGATCCCGGCTTCCACTTCGAGGACACCGCATCGAGATC 606
DB 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgileGluile 201
QY 607 CTGAGGAGGTGGAGAGGCGAAGTGTACAAGCAGTACGAGCCCGCTGGCGCGCTAC 666
DB 202 MetGluGluValGluLysGlyLysCysTyrgLysGlnTyrgLysAlaValGlyArgTyrg 221
QY 667 TGCACGCGCGCCCTCCAACTGGGCGCAAC 699
DB 222 CysaspAlaAlaProSerLysLeuGlyHisasn 232

```

RESULT 13

ID AAE28834

XX AAE28834 standard; protein; 232 AA.

AC AAE28834;

XX 27-DEC-2002 (first entry)

DE Anemonia sulcata asP600 (NFP-7) wild-type protein.

XX

KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
 KW fluorescence activated cell sorting application; fluorescent timer;
 KW biosensor; fluorescence resonance energy transfer application; FRET;
 KW colouring agent; recombinant DNA application; analyte detection assay;
 KW sunscreen; second messenger detector; asP600 protein; NFP-7.
 XX Anemonia sulcata.
 OS
 XX Key Location/Qualifiers
 FH Misc-difference 68
 FT /note= "Encoded by GCC"
 FT Misc-difference 143
 FT /note= "Encoded by TCC"
 XX
 XX WO200268459-A2.
 XX 06-SEP-2002.
 XX 20-FEB-2002; 2002WO-US005749.
 XX 21-FEB-2001; 2001US-0270983P.
 XX 04-DEC-2001; 2001US-00006922.
 XX (CLON-) CLONTECH LAB INC.
 PA Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Pradkov A;
 XX WPI; 2002-691654/74.
 XX N-PSDB; AAD46279.
 XX
 XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
 PT of an aggregating Chidarian chromo- or fluorescent protein or mutant for
 PT analyte detection assays or fluorescence activated cell sorting
 PT applications.
 XX Disclosure; Page 71-72; 80pp; English.

PS The invention relates to nucleic acid molecules encoding non-aggregating
 XX chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
 CC useful in analyte detection assays, as colouring agents, as markers in
 CC recombinant DNA applications, as sunscreens or filters, in fluorescence
 CC resonance energy transfer (FRET) applications, as biosensors in
 CC prokaryotic and eukaryotic cells, in screening assays, as second
 CC messenger detectors, in fluorescence activated cell sorting applications,
 CC in protease cleavage assays or as fluorescent timers. The present
 CC sequence is Anemonia sulcata asP600 (NFP-7) wild-type protein of the
 CC invention

SQ Sequence 232 AA;

Alignment Scores:

Pred. No.: 1-25e-86 Length: 232
 Score: 1216.00 Matches: 220
 Percent Similarity: 97.84% Conservatives: 6
 Best Local Similarity: 95.24% Mismatches: 5
 Query Match: 93.54% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-21 (1-707) x AAE28834 (1-232)

```

QY 7 GCCTCCCTGTCAGCAGACACATCCCTTCAGACACACATCCAGGCGCCGCTGAACGCG 66
DB 2 AlaserPheLeuLysThrMetProPheLysThrThrileGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTCCACCGGCAAGGCGAGGCAACCCCTCGAGGCGCACCCAGGAGATG 126
DB 22 Histy-PheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCGAGGTATCGAGGCGGCGCCCTCGCTTCCCTTCACATCTCGTCCACTCC 186
DB 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCGGCATCCCGCACTACTTCAAG 246

```

Db 62 CysMetTyrGlySerLysThrPheLeuLysTyrValSerGlyIleProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGGCTTCACTGGGAGCGCCACCACTACGAGACGGCGGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCGACAGGACACTCCCTGACGGGAGCTGCTGGTGTACAAGGTGAAGATC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
 QY 367 CTGGCAACAACCTTCCCGCGAGCGCCCGTGTGATGAGAACAGGCGCGCTGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyA-grpGlu 141
 QY 427 CCTCCACCGAGATCGTGTACGAGGTGAGCGGCGTGTGCGCGCCAGTCCCTGATGGCC 486
 Db 142 ProAlaThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGAGTGCCTCCGCGGTCGACCTGACCTGCGACCTGCGACCTGCGACCTACCGCTCAAG 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCCGCGCTCCGCGCTGAAGATGCCGCGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
 Db 182 LysProAlaAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
 QY 607 CTGAGGAGGTGGAGAGGCAAGTGTACAGCAGTACAGGCGCGCGCTGGCGCCCTAC 666
 Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
 QY 667 TGCAGCCGCGCGCTCCCAAGCTGGGCGCAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 14

AAG79764

ID AAG79764 standard; protein; 232 AA.

XX AC AAG79764;

XX DT 01-APR-2003 (first entry)

XX DE Kindling fluorescent protein, AsFP595 F90L, A148G, H203Y.

XX KW Kindling fluorescent protein; kindling stimulus; movement; labeling;

XX KW fluorescence resonance energy transfer; FRET;

XX KW bioluminescence resonance energy transfer; BRET; biosensor;

XX KW automated screening.

XX OS Anthozoa sulcata.

XX FH Key Location/Qualifiers

FT Misc-difference 90

FT FT /label= F90L

FT FT Misc-difference 148

FT FT /label= A148G

FT FT Misc-difference 203

FT FT /label= H203Y

XX PN WO200296924-A1.

XX PD 05-DEC-2002.

XX XX 24-MAY-2002; 2002WO-US016379.

XX XX 25-MAY-2001; 2001US-0293752P.

XX XX 11-OCT-2001; 2001US-0329176P.

XX XX (CLON-) CLONTECH LAB INC.

XX PA Lukyanov SA, Chudakov D, Lukyanov K;

XX PI

XX XX

DR

DR

XX

PT

PT

XX

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Sequence 232 AA;

Alignment Scores:

Pred. No.: 7,59e-86 Length: 232

Score: 1206.00 Matches: 219

Percent Similarity: 96.97% Conservative: 5

Best Local Similarity: 94.81% Mismatches: 7

Query Match: 92.77% Indels: 0

DB: 6 Gaps: 0

US-10-081-864-21 (1-707) x AAG79764 (1-232)

QY 7 GCCTCCCTGTGACCGAGACCATGCCCTTCAGGACCAACATCGAGGCGCCGTTGAACGGC 66

Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21

QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGGCAACCCCTCGAGGCGCCACCGAGGAGATG 126

Db 22 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41

QY 127 AAGATCGAGTGTATCGAGGCGCGCCCTTCGCTTCACATCTCTGTCACCTCC 186

Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61

QY 187 TGCATGTACGGCTCCCAAGCGCTTCATCAAGTGTGCGGCGATCCCGACTACTTCAAG 246

Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 81

QY 247 CAGTCCCTCCCGAGGGCTTCACTGGGAGCGCCACCACTACGAGACCGCGGCTTC 306

Db 82 GlnSerPheProGluGlyLeuThrTrpGluArgThrThrThrTyrGluAspGlyGlyPhe 101

QY 307 CTGACCGCCGACGAGACACTCCCTGGAGCGGCGACTGCTGCTGTACAGGTGAAGATC 366

Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121

QY 367 CTGGCAACAACCTTCCCGCGAGCGCCCGTGTGAGAACAGGCGCGCTGGAG 426

Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141

QY 427 CCCTCCACCGAGATCGTGACGAGGTGGACGGGTGCTGCGCGCCAGTCCCTGTATGGCC 486
Db ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGAGTCCCGCGGGTCCGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 546
Db LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCCCGCTCCGCGCTGAGATCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
Db LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspTyrArgIleGluIle 201
QY 607 CTGAGGAGGTGGAGAGGCAAGTGTACACAGTACAGGACGAGGCGCGCTGCGCTAC 666
Db MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
QY 667 TGGACGCGCGCGCTCCCAAGCTGGGCCACAAC 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 15

AAG79467

ID AAG79467 standard; protein; 226 AA.

XX AC AAG79467;

XX DT 29-NOV-2002 (first entry)

XX DE C. gigantea mutant chromoprotein.

XX KW Chromoprotein; CgCP; C. gigantea; fluorescent; emission maximum;
XX KW labeling; colouring agent; food; pharmaceutical; cosmetic; biosensor;
XX KW fluorescence resonance energy transfer; detection assay; FRET;
XX KW selectable marker.

XX OS Condylactis gigantea.

XX PH Key Location/Qualifiers

FT Misc-difference 142

FT /label= C148S

FT /note= "Numbering according to GFP"

XX WO200259309-A2.

XX PD 01-AUG-2002.

XX PF 11-DEC-2001; 2001WO-US047995.

XX PR 13-DEC-2000; 2000US-0255533P.

XX PA (CLON-) CLONTECH LAB INC.

XX PI Lukyanov SA, Lukyanov KA, Fradkov AF;

XX DR WPI; 2002-666902/71.

XX DR N-PSDB; ABA00245.

XX PT New nucleic acids encoding chromoproteins or fluorescent proteins, useful
XX PT as labeling tools for marking a protein, cell or organism, which may be
XX PT used in biochemistry, molecular biology or medical diagnostic
XX PT applications.

XX PS Disclosure; Fig 2; 56pp; English.

XX CC The sequences given in AAG79466-67 represent wild type and mutant
XX CC chromoprotein (CgCP) from C. gigantea. The mutant protein is fluorescent
XX CC having an emission maximum ranging from about 480-680 nm. The wild type
XX CC and mutant proteins are useful as labeling tools for marking a protein,
XX CC cell or organism, which may be used in biochemistry, molecular biology or
XX CC medical diagnostic applications. The chromoproteins may be used as
XX CC colouring agents capable of imparting colour or pigment to a particular
XX CC composition of matter, e.g. food compositions, pharmaceuticals or
XX CC cosmetics. They may also be used as labels in analyte detection assays,

CC or as selectable markers in recombinant DNA applications. The fluorescent
CC proteins may be used in fluorescence resonance energy transfer (FRET)
CC applications, e.g. detection of protein-protein interactions. They are
CC also used as biosensors in prokaryotic and eukaryotic cells, e.g. as pH
CC indicator or phosphorylation indicator. The proteins are useful as in
CC vivo marker in animals, in assays to determine the phospholipid
CC composition in biological membranes, or in protease cleavage assays

XX SQ Sequence 226 AA;

Alignment Scores: 4.64e-53 Length: 226
Pred. No.: 786.50 Matches: 149
Score: 78.85% Conservative: 30
Percent Similarity: 78.85%
Best Local Similarity: 65.64% Mismatches: 45
Query Match: 60.50% Indels: 3
DB: 5 Gaps: 3

US-10-081-864-21 (1-707) x AAG79467 (1-226)

QY 7 GCCTCCCTGTCACCGAGACCATGCCCTTCAGGACCACTCAGAGGCGACCGTGAACGCG 66
Db 1 AlaGlyLeuLeuLysGluSerMetArgIleLysIleTyrMetGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGCGACCCAGGAGATG 126
Db 21 TyrHisPheLysCysGluGlyGluGlyAspGlyAsnPropheGluGlyThrGlnAsnMet 40
QY 127 AAGATCGAGGTGATCGAGGGGCGCCCTCGCTTCCCTCCACATCGTCCACCTCC 186
Db 41 ArgIleArgValThrGluGlyAlaProLeuPropheAlaPheAspIleLeuSerProCys 60
QY 187 TGCATGTACGCGCTCCAAAGGCTTCATCAAGTACGTGCTCGGCGATCCCGGACTACTTCAAG 246
Db 61 CysAlaTyrGlySerLysThrPheIleLysHisThrSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCTCCCGGAGGCTTCACTGGGAGCGCACCACTACGAGGACGCGGCTTC 306
Db 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrIleTyrGluAspGlyGlyVal 100
QY 307 CTGACCGCGCCACGAGGACACCTCCCTGGAGCGGACTGCTGTGTACAGGTGAAGATC 366
Db 101 LeuThrAlaHisGlnAspThrSerLeuGluGlyAsnCysLeuIleTyrLysVal 120
QY 367 CTGGGCAACACTTCCCGCGGCGCGCTGATGCAGAACAGGCGCGCGCTGGAG 426
Db 121 LeuGlyThrAsnPheProAlaAspGlyProValMetLysLysIleSerGlyGlyTrpGlu 140
QY 427 CCCTCCACCGAGATCGGTACGAGGTGGACGGCGTGTGCTGCGCGCGCAGTCCCTGATGGCC 486
Db 141 ProSerThrGluIleValTyrGlnAspAsnGlyValLeuArgGlyArgAsnValMetAla 160
QY 487 CTGAGTGGCGCGCGCTCGCACCTGACCTGCACCTGCACACCACTACCGTCCAG 546
Db 161 LeuLysValSerGlyArgProProLeuIleCysHisLeuHisSerThrTyrArgSerLys 180
QY 547 AAGCCCGCTCCGCGCTGAGATCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
Db 181 Lys---AlaCysAlaLeuThrMetProGlyPheHisPheAlaAspLeuArgIleGln--- 198
QY 607 CTGAGGAGGTGGAGAGGCAAGTGTACAGCAGTACAGGCGCGCGCTGGCGCGCTAC 666
Db 199 MetProLysLysLysLysAspGluTyrPheGluLeuTyrGluAlaSerValAlaArgTyr 218
QY 667 TGGACGCGCGCGCTCCCAAG 687
Db 219 SerAsp---ValProGluLys 224

Search completed: July 29, 2004, 14:32:30
Job time : 54.2812 secs

Blank Sheet USPTO

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:29:01 ; Search time 14.5484 Seconds
(without alignments)
5017.682 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 1300

Sequence: 1 ggatccgcctcctcgtgac.....ctgggcacacactgaagctt 707

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10081864/runat_23072004_150507_25519/app_query.fasta_1.1742
-DB=Issued Patents AA -Qfmt=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USMR=US10081864@cgn_1.1_45@runat_23072004_150507_25519 -RCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pcp.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pcp.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pcp.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pcp.*
- 5: /cgn2_6/ptodata/2/iaa/6C_COMB.pcp.*
- 6: /cgn2_6/ptodata/2/iaa/6D_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	453.5	34.9	238	3	US-09-277-716-32
2	453.5	34.9	238	4	US-09-609-161B-32
3	445.5	34.3	238	3	US-09-277-716-16
4	445.5	34.3	238	4	US-09-609-161B-16
5	445.5	34.3	238	4	US-09-626-581B-65
6	445.5	34.3	238	4	US-09-415-765B-65
7	445.5	34.3	238	4	US-09-626-580C-65
8	433.5	33.3	238	4	US-09-839-650-3
9	239	17.8	377	4	US-09-252-991A-31793
10	232.5	17.9	247	4	US-09-252-991A-26899
11	228.5	17.6	238	4	US-09-472-065A-5
12	228.5	17.6	997	4	US-09-417-197-121

13	227	17.5	302	4	US-09-252-991A-19529	Sequence 19529, A
14	225.5	17.3	631	4	US-09-417-197-39	Sequence 39, Appl
15	225	17.3	322	4	US-09-252-991A-31608	Sequence 31608, A
16	224.5	17.3	238	4	US-09-023-946B-20	Sequence 20, Appl
17	224.5	17.3	238	4	US-09-023-946B-30	Sequence 30, Appl
18	224.5	17.3	238	4	US-09-023-946B-31	Sequence 31, Appl
19	224.5	17.3	238	4	US-09-023-946B-32	Sequence 32, Appl
20	224.5	17.3	238	4	US-09-023-946B-33	Sequence 33, Appl
21	224.5	17.3	238	4	US-09-023-946B-37	Sequence 37, Appl
22	224.5	17.3	1070	4	US-09-091-042A-2	Sequence 2, Appl
23	224	17.2	843	4	US-09-417-197-117	Sequence 117, App
24	223.5	17.2	238	3	US-08-643-704A-49	Sequence 49, Appl
25	223.5	17.2	242	4	US-09-023-946B-1	Sequence 1, Appl
26	222.5	17.1	218	4	US-09-603-448-5	Sequence 5, Appl
27	222.5	17.1	238	1	US-08-452-295-1	Sequence 1, Appl
28	222.5	17.1	238	4	US-09-213-343-4	Sequence 4, Appl
29	222	16.6	197	4	US-09-252-991A-26537	Sequence 26537, A
30	222	17.1	1171	4	US-09-417-197-131	Sequence 131, App
31	220.5	17.0	239	3	US-09-094-359-10	Sequence 10, Appl
32	220.5	17.0	239	3	US-09-172-083-5	Sequence 5, Appl
33	220.5	17.0	239	4	US-09-316-919-6	Sequence 6, Appl
34	220.5	17.0	239	4	US-09-602-641-5	Sequence 5, Appl
35	220.5	17.0	239	4	US-09-704-463-10	Sequence 10, Appl
36	219.5	16.9	238	4	US-09-023-946B-24	Sequence 24, Appl
37	219	16.8	941	4	US-09-513-783A-172	Sequence 172, App
38	218.5	16.8	238	3	US-08-893-327-16	Sequence 16, Appl
39	218.5	16.8	238	4	US-09-472-085A-6	Sequence 6, Appl
40	218.5	16.8	239	3	US-09-094-359-4	Sequence 4, Appl
41	218.5	16.8	239	3	US-09-172-063-3	Sequence 3, Appl
42	218.5	16.8	239	4	US-09-513-783A-46	Sequence 46, Appl
43	218.5	16.8	239	4	US-09-316-919-4	Sequence 4, Appl
44	218.5	16.8	239	4	US-09-602-641-3	Sequence 3, Appl
45	218.5	16.8	239	4	US-09-704-463-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-277-716-32
; Sequence 32, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLINE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT FILING DATE: 1999-03-26
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; OTHER INFORMATION: Ptilosarcus gurneyi Green Fluorescent Protein (GFP)
US-09-277-716-32

Alignment Scores:
Pred. No.: 6,548-29 Length: 238
Score: 453.50 Matches: 94
Percent Similarity: 58.33% Conservative: 32
Best Local Similarity: 43.52% Mismatches: 87
Query Match: 34.88% Indels: 3
DB: 3 Gaps: 2

US-10-081-864-21 (1-707) x US-09-277-716-32 (1-238)

16 CTGACCGAGACCATGCTCCCTCAGGACACCATCAGGAGGACCGTGAAGCGCCACTACTTC 75
11 LeuLysGluIleMetSerAlaLysAlaSerValGluGlyIleValAsnHisValPhe 30
76 AAGTGCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 135
31 SerMetGluGlyPheGlyLysGlyAsnValLeuPheGlyAsnGlnLeuMetGlnIleArg 50
136 GTGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 195
51 ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIleAlaPheGlnTyr 70
196 GAGTCCCAAGGCGCTTCATCAAGTACGTGTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 255
71 GlyAsnArgThrPheThrLysTyrProAspIleAlaAspTyrPheValGlnSerPhe 90
256 CCGGAGGCGCTTCACCTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 315
91 ProAlaGlyPhePheTyrGluArgAsnLeuArgPheGluAspGlyAlaIleValAspIle 110
316 CACGAGGACACTCCCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 375
111 ArgSerAspIleSerLeuGluAspAspLysPheHisTyrLysValGluTyrArgGlyAsn 130
376 AACTTCCCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
131 GlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMetGluProSerPhe 150
436 GAGATCGTGTACGAGTGCACCTGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495
171 GluSerGlyAsnTyrSerCysHisMetLysThrPheTyrArgSerLys-----Gly 198
556 TCCGCGCTGAAGTCCCGGCTTCACCTTCAGGACACCGCATCGAGTCCCTGGAGGAG 615
189 GlyValLysGluPheProGluTyrHisPheIleHisArgLeuGlu-----LysThrTyr 207
616 GTGAGAGGCGCAAGTGTCTACAGAGTACAGAGTACAGAGGCGGCGGCGGCGGCGGCGG 663
208 ValGluGluGlySerPheValGluGlnHisGluThrAlaIleAlaGln 223

RESULT 2

US-09-609-161B-32
; Sequence 32, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUNE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HI
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi

FEATURE:
; OTHER INFORMATION: Ptilosarcus gurneyi Green Fluorescent Protein (GFP)
US-09-609-161B-32

Alignment Scores:
Pred. No.: 6,548-29 Length: 238
Score: 453.50 Matches: 94
Percent Similarity: 58.33% Conservative: 32
Best Local Similarity: 43.53% Mismatches: 87
Query Match: 34.88% Indels: 3
DB: 4 Gaps: 2

US-10-081-864-21 (1-707) x US-09-609-161B-32 (1-238)

QY 16 CTGACCGAGACCATGCTCCCTCAGGACACCATCAGGAGGACCGTGAAGCGCCACTACTTC 75
Db 11 LeuLysGluIleMetSerAlaLysAlaSerValGluGlyIleValAsnHisValPhe 30
QY 76 AAGTGCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 135
Db 31 SerMetGluGlyPheGlyLysGlyAsnValLeuPheGlyAsnGlnLeuMetGlnIleArg 50
QY 136 GTGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 195
Db 51 ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIleAlaPheGlnTyr 70
QY 196 GAGTCCCAAGGCGCTTCATCAAGTACGTGTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 255
Db 71 GlyAsnArgThrPheThrLysTyrProAspIleAlaAspTyrPheValGlnSerPhe 90
QY 256 CCGGAGGCGCTTCACCTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 315
Db 91 ProAlaGlyPhePheTyrGluArgAsnLeuArgPheGluAspGlyAlaIleValAspIle 110
QY 316 CACGAGGACACTCCCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 375
Db 111 ArgSerAspIleSerLeuGluAspAspLysPheHisTyrLysValGluTyrArgGlyAsn 130
QY 376 AACTTCCCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
Db 131 GlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMetGluProSerPhe 150
QY 436 GAGATCGTGTACGAGTGCACCTGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495
Db 151 GluValValTyrMetAsnSerGlyValLeuValGlyGluValAspLeuValTyrLysLeu 170
QY 496 CCGGCGGCTGCACCTGACCTGCCACCTGCAGACACCGCATCGAGTCCCTGGAGGAG 555
Db 171 GluSerGlyAsnTyrSerCysHisMetLysThrPheTyrArgSerLys-----Gly 188
QY 556 TCCGCGCTGAAGTCCCGGCTTCACCTTCAGGACACCGCATCGAGTCCCTGGAGGAG 615
Db 189 GlyValLysGluPheProGluTyrHisPheIleHisArgLeuGlu-----LysThrTyr 207
QY 616 GTGAGAGGCGCAAGTGTCTACAGAGTACAGAGTACAGAGGCGGCGGCGGCGGCGGCGG 663
Db 208 ValGluGluGlySerPheValGluGlnHisGluThrAlaIleAlaGln 223

RESULT 3

US-09-277-716-16
; Sequence 16, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUNE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15

EARLIER APPLICATION NUMBER: 60/079,624

EARLIER FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 16

LENGTH: 238

TYPE: PRT

ORGANISM: Renilla mulleri

FEATURE:

OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)

US-09-277-716-16

Alignment Scores:
Pred. No.: 238
Score: 445.50
Percent Similarity: 59.26%
Best Local Similarity: 42.55%
Query Match: 34.27%
DB: 3

US-10-081-864-21 (1-707) x US-09-277-716-16 (1-238)

```
QY 16 CTGACCGAGACCATCCCTTCAGGACACCATCGAGGCGACCGTCAAGCGCCACTACTTC 75
DB 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyIleValAsnHisValPhe 30
QY 76 AAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATCGAG 135
DB 31 ThrMetGluGlyCysGlyLysGlyAsnLeuPheGlyAsnGlnLeuValGlnIleArg 50
QY 136 GTGATCGAGGGCGGCGGCTCGCTCCACATCCCTCCACCTCCCTCCCTCCCTCCCTGATC 195
DB 51 ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
QY 196 GGTCTCAAGCGCTTCATCAAGTACGTCTCGGCGATCCCGACTCTTCAAGCATGCTCCTC 255
DB 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
QY 256 CCGAGGGCTTCACTGAGGCGGCGGCTCGGCGGCGGCGGCTCGGCGGCGGCGGCGGCGG 315
DB 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
QY 316 CACGAGGACCTCCCTCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 375
DB 111 ArgSerAspIleAsnLeuLeuGluValGluValGluValGluValGluValGluVal 130
QY 376 AACTTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
DB 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
QY 436 GAGATCGTGTACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 495
DB 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGluValGluValGluValGluVal 170
QY 496 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555
DB 171 AsnSerGlyLysTyrSerCysHisMetLysThrLysThrLysThrLysThrLysThr 188
QY 556 TCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 615
DB 189 ValValLysGluPheProSerTyrHisPheIleGlnHisArgLeuGlu---LysThrTyr 207
QY 616 GTGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 663
DB 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223
```

RESULT 4

US-09-609-161B-16

Sequence 16, Application US/09609161B

Patent No. 6436682

GENERAL INFORMATION:

APPLICANT: Bryan, Bruce

APPLICANT: Szent-Gyorgyi, Christopher

APPLICANT: PROLUME, LTD.

TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC

TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG

FILE REFERENCE: 24729-121B

CURRENT APPLICATION NUMBER: US/09/609,161B

CURRENT FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: 09/277,716

PRIOR FILING DATE: 1999-03-26

PRIOR APPLICATION NUMBER: 60/102,939

PRIOR FILING DATE: 1998-10-01

PRIOR APPLICATION NUMBER: 60/089,367

PRIOR FILING DATE: 1998-06-15

PRIOR APPLICATION NUMBER: 60/079,624

PRIOR FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 16

LENGTH: 238

TYPE: PRT

ORGANISM: Renilla mulleri

FEATURE:

OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)

US-09-277-716-16

Alignment Scores:
Pred. No.: 238
Score: 445.50
Percent Similarity: 59.26%
Best Local Similarity: 42.55%
Query Match: 34.27%
DB: 3

US-10-081-864-21 (1-707) x US-09-277-716-16 (1-238)

```
QY 16 CTGACCGAGACCATCCCTTCAGGACACCATCGAGGCGACCGTCAAGCGCCACTACTTC 75
DB 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyIleValAsnHisValPhe 30
QY 76 AAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATCGAG 135
DB 31 ThrMetGluGlyCysGlyLysGlyAsnLeuPheGlyAsnGlnLeuValGlnIleArg 50
QY 136 GTGATCGAGGGCGGCGGCTCGCTCCACATCCCTCCACCTCCCTCCCTCCCTGATC 195
DB 51 ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
QY 196 GGTCTCAAGCGCTTCATCAAGTACGTCTCGGCGATCCCGACTCTTCAAGCATGCTCCTC 255
DB 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
QY 256 CCGAGGGCTTCACTGAGGCGGCGGCTCGGCGGCGGCGGCTCGGCGGCGGCGGCGGCGG 315
DB 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
QY 316 CACGAGGACCTCCCTCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 375
DB 111 ArgSerAspIleAsnLeuLeuGluValGluValGluValGluValGluValGluVal 130
QY 376 AACTTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
DB 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
QY 436 GAGATCGTGTACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 495
DB 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGluValGluValGluValGluVal 170
QY 496 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555
DB 171 AsnSerGlyLysTyrSerCysHisMetLysThrLysThrLysThrLysThrLysThr 188
QY 556 TCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 615
DB 189 ValValLysGluPheProSerTyrHisPheIleGlnHisArgLeuGlu---LysThrTyr 207
QY 616 GTGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 663
DB 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223
```

Db 189 ValVallysGluPheProSerTyrHisPheIleGlnHisArgLeuGlu---LysThrTyr 207
QY 616 GTGAGAGAGGCAAGTGTACAGCAGTACAGAGCCCGCTGGCGCGC 663
Db 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223
RESULT 5
US-09-626-581D-65
; Sequence 65, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-626-581D-65

Alignment Scores:
Pred. No.: 2,9e-28 Length: 238
Score: 445.50 Matches: 92
Percent Similarity: 59.26% Conservatives: 36
Best Local Similarity: 42.59% Mismatches: 85
Query Match: 34.27% Indels: 3
Gaps: 2

US-10-081-864-21 (1-707) x US-09-626-581D-65 (1-238)

QY 16 CTGACCGAGACCATGCGCTTCAGGACCAACCATCGAGGACCGCTGAACGGCCACTCTTC 75
Db 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyLeuValAsnHisValPhe 30
QY 76 AAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGCGCACCGAGATGAAGATCGAG 135
Db 31 ThrMetGluGlyCysGlyLysGlyAsnIleLeuPheGlyAsnGlnLeuValGlnIleArg 50
QY 136 GTGATCGAGGCGGCGCCCTCGCTTCACATCTGTGCTTCCATCTGCTCCACTCTCGCATGTAC 195
Db 51 ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
QY 196 GGCTCCAAAGCGCTTCATCAAGTACGTGTCGGGATCCCGGATCCCGACTACTTCAAGCAGTCCCTC 255
Db 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
QY 256 CCGAGGCGCTTACCTCGGAGCGCACCAACCATCGAGGAGCGGCGCTTCCTGACCGCC 315
Db 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
QY 316 CACCAGGACACCTCCCTGGAGCGGCGCTGCTGTGTACAAAGTGAAGATCTCTGGGCAAC 375
Db 111 ArgSerAspIleAsnLeuIleGluAspPheValTyrArgValGluTyrLysGlySer 130
QY 376 AACTTCCCGCGACCGCCGCTGTGTACAGAACAAAGCGCGCGCTGGGAGCGCTCCACCC 435
Db 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
QY 436 GAGATCGTACGAGTGAACGCGCTGCTGGCGGCGCAGTCCCTGATGCGCCCTGGAGTGC 495
Db 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyValIleLeuValTyrLysLeu 170
QY 496 CCGCGCGGTGGCCACCTGACCTGCCACCTGCACACCATCCGCTCCAAAGAGCCCGCC 555

Db 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLys-----Gly 188
QY 556 TCCGCGCTGAGATGCCCGGCTTCACACTTCGAGGACCAACCGCATCGAGATCCTGGAGGAG 615
Db 189 ValVallysGluPheProSerTyrHisPheIleGlnHisArgLeuGlu---LysThrTyr 207
QY 616 GTGAGAGAGGCAAGTGTCTACAAGCAGTACAGGCGCGCTGGCGCGC 663
Db 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223

RESULT 6

US-09-415-765B-65
; Sequence 65, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-415-765B-65

Alignment Scores:
Pred. No.: 2,9e-28 Length: 238
Score: 445.50 Matches: 92
Percent Similarity: 59.26% Conservatives: 36
Best Local Similarity: 42.59% Mismatches: 85
Query Match: 34.27% Indels: 3
Gaps: 2

US-10-081-864-21 (1-707) x US-09-415-765B-65 (1-238)

QY 16 CTGACCGAGACCATGCGCTTCAGGACCAACCATCGAGGCGCACCGTGAACGGCCACTCTTC 75
Db 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyLeuValAsnHisValPhe 30
QY 76 AAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGCGCACCGAGATGAAGATCGAG 135
Db 31 ThrMetGluGlyCysGlyLysGlyAsnIleLeuPheGlyAsnGlnLeuValGlnIleArg 50
QY 136 GTGATCGAGGCGGCGCCCTCGCTTCACATCTGTGCTTCCATCTGCTCCACTCTCGCATGTAC 195
Db 51 ValThrLysGlyAlaProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
QY 196 GGCTCCAAAGCGCTTCATCAAGTACGTGTCGGGATCCCGGATCCCGACTACTTCAAGCAGTCCCTC 255
Db 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
QY 256 CCGAGGCGCTTACCTCGGAGCGCACCAACCATCGAGGAGCGGCGCTTCCTGACCGCC 315
Db 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
QY 316 CACCAGGACACCTCCCTGGAGCGGCGCTGCTGTGTACAAAGTGAAGATCTCTGGGCAAC 375
Db 111 ArgSerAspIleAsnLeuIleGluAspPheValTyrArgValGluTyrLysGlySer 130
QY 376 AACTTCCCGCGACCGCCGCTGTGTACAGAACAAAGCGCGCGCTGGGAGCGCTCCACCC 435
Db 131 AsnPheProAspAspGlyProValMetGlnLysThrIleLeuGlyIleGluProSerPhe 150
QY 436 GAGATCGTACGAGTGAACGCGCTGCTGGCGGCGCAGTCCCTGATGCGCCCTGGAGTGC 495
Db 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyValIleLeuValTyrLysLeu 170

Qy 436 GAGATCGTGTCAGAGGTGGAGCGCGTGGCGCGGCAGCTCCCTGATGGCCCTGGAGTGC 495
 Db 151 GluAlaMetTyrMetAsnAsnGlyValIleuValGlyGluValIleuValTyrLysLeu 170
 Qy 496 CCGCGCGTGGCACCTGACCTGCACCTGCACACCTACCGCTCCAGAGCCCGCC 555
 Db 171 AnnSerGlyLysTyrTyrSerCysHisMetLysThrLieuMetLysSerLys-----Gly 188
 Qy 556 TCGCGCCTAGATCCCGCGGTCTCCATCTTGAGAGCACCGCATCAGATCTCTGGAGGAG 615
 Db 189 ValValLysGluPheProSerTyrHisPheIleGlnHisArgLeuGlu---LysThrTyr 207
 Qy 616 GTGGAGAGGGCAAGTGTCTAAGCAGGTACGAGCGCCCGCTGGCGCGC 663
 Db 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223

RESULT 8
 US-09-839-650-3
 ; Sequence 3, Application US/09839650
 ; Patent No. 6645761
 ; GENERAL INFORMATION:
 ; APPLICANT: Stratagene
 ; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mull
 ; Patent No. 6645761
 ; TITLE OF INVENTION: Fluorescent Protein
 ; FILE REFERENCE: 25436/1755
 ; CURRENT APPLICATION NUMBER: US/09/839,650
 ; CURRENT FILING DATE: 2001-04-19
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 238
 ; TYPE: PRP
 ; ORGANISM: Renilla muelleri

```

Alignment Scores:
Pred. No.:      2,71e-27      Length:      238
Score:          433.50      Matches:      91
Percent Similarity: 58.33%      Conserves:      35
Best Local Similarity: 42.13%      Mismatches:      87
Query Match:      33.35%      Indels:      3
DB:              4              Gaps:      2

US-10-081-864-21 (1-707) x US-09-839-650-3 (1-238)

Qy 16 CTGACGAGACCATGCCCTTCAGGACCAACCATCGAGGGCACCGTGAAAGCGGCACACTCTTC 75
Db 11 LeuGlnGluValMetSerTyrLysValAsnLeuGluGlyLeValAsnAsnHisValPhe 30
Qy 76 AAGTGCACCGCAAGGGGCGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATCGAG 135
Db 31 ThrMetGluGlyCysArgLysArgAsnIleLeuPheGlyAsnGlnLeuValHisIleArg 50
Qy 136 GTGATCGAGGGCGGCCCTCGCTTCGCTTCACATCTCTGTCCACCTCTCTCGATGTAC 195
Db 51 ValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerProAlaPheGlnTyr 70
Qy 196 GGCCTCAAGCGCTTCATCAAGTAGTGTCGGCGATCCCCGACTACTTCAGCAGAGTCCCTC 255
Db 71 GlyAsnArgThrPheThrLysTyrProAsnAspIleSerAspTyrPheIleGlnSerPhe 90
Qy 256 CCGAGGGGTTCCCTCGGAGCGCACCAACCTACGAGACCGGGCGGCTTCCTGACCGGC 315
Db 91 ProAlaGlyPheMetTyrGluArgThrLeuArgTyrGluAspGlyGlyLeuValGluIle 110
Qy 316 CACCAAGACACCTCCCTCGAGGGGACCTGCTGTGTACAAAGGTGAAGATCCTGGGCAAC 375
Db 111 ArgSerAspIleAsnLeuIleGluAspLysPheValTyrArgValGluTyrLysGlySer 130
Qy 376 AACTTCCCGCGACGGCCCGCTGTATGCAGAAACGAGCGCGCGCTGGGAGCCCTCCACC 435

```

```

Db 131 AsnPheProAspGlyProValMetGlnLysThrLeuGlyLeuGluProSerPhe 150
QY 436 GAGATCGGTACGAGGTGGCGGTGCTGGCGCGGCGAGTCCCTGATGGCCCTGGAGTGC 495
Db 151 GluAlaMetTyrMetAsnAsnGlyValLeuValGlyGluValIleLeuValTyrLysLeu 170
QY 496 CCCGGCGTCCACCTGACCTGACCTGCCACCTGCACACCACTACCGCTCCAGAGACCGGCC 555
Db 171 AsnSerGlyLysTyrTyrSerCysHisMetLysThrLeuMetLysSerLys-----Gly 188
QY 556 TCCGCCCTGAGATGCCCGGTTCCTACTTCGAGGACCGCATCGAGATCCTGGAGGAG 615
Db 189 ValValLysGluPheProSerTyrHisPheIleGlnHisArgLeuGlu-----LysThrTyr 207
QY 616 GTGGAGAGGGCAAGTGTACAGCAGTACGAGCGCCCGTGGCGCGC 663
Db 208 ValGluAspGlyGlyPheValGluGlnHisGluThrAlaIleAlaGln 223

RESULT 9
US-09-252-991A-31793
; Sequence 31793, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31793
; LENGTH: 377
; TYPE: PPT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (59)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-31793

Alignment Scores:
Pred. No.: 1.64e-11 Length: 377
Score: 239.00 Matches: 96
Percent Similarity: 34.86% Conservative: 18
Best Local Similarity: 29.36% Mismatches: 89
Query Match: 17.82% Indels: 124
DB: 4 Gaps: 20

US-10-081-864-21 (1-707) x US-09-252-991A-31793 (1-377)
QY 700 AGTTGTGCCAGCTGTGAGGGGCGG----- 674
Db 37 AlaCysAlaProAlaGlyAlaGlyArgGlyTyrLeuProAlaArgPheArgArg 56
QY 673 -----CGTCCGACGT----- 665
Db 57 AlaPro***AlaGlyArgArgGlyArgProAlaIleArgSerAlaValAlaArg 76
QY 664 ---AGCGGCCACCGCGCCCT-----CGTACTGCTTGT 635
Db 77 ArgAlaGlyAlaArgArgProGlySerHisGlyArgArgSerAlaGlyArgValAlaAla 96
QY 634 AGCACTTCCCTTCCACCTCCCTCCAGATCTCGATCGGTGGT-----CCTCGA 584
Db 97 GlyArgAlaProSer---AlaAlaProGlySerArgIleGlyAlaValArgProAla 115
QY 583 AGTGAAGCCGGCATCTTCAGGCGGAGCGGGGCTTCTTGGACGGTAGTGTGTGCA 524
Db 116 AlaAlaProArgCysArgAsnAlaProArgAlaGly-----ProCysGlu 131

```

```

QY 523 GGTGCGAG-----TCAGTGGCGACCGCGCGGGGCACTCCAGGG 485
Db 132 ArgGlyArgAlaThrProProArgAspAlaAlaGlyTyrArgArgSerSerAlaPro 151
QY 484 CCATCAGGAGTGGCGCGGCGACCGCTCCCTCCACCATCTCGGTGGAGGGCT 425
Db 152 AlaAlaAlaGlyArgSerArgAlaProProValArgAspThrTyrProAla 171
QY 424 CCC---AGCGGCGCGCTTGTCTCATCACGGGCGCTCGCGGGGAAGT----- 377
Db 172 ProValArgGlyArgSerArgAlaArgAlaArgAlaThrGlyArgPro 191
QY 376 -----TCTTGGCCAGGATCTTCACCTTGT 353
Db 192 HisTyrLeuProValArgAsnAlaAlaArgAlaIleArgCysCysProAlaThrAlaProCys 211
QY 352 ACACCGAGCGAGTCGC-----CGTCCAGGAGGTGTCTCTGGTGGCGGTCA 308
Db 212 ValProGlyHisArgAspAlaArgArgGlyAlaProGlyArg---ProGlySerArgAla 230
QY 307 GGAAGC-----CGCGTCTCT 293
Db 231 GlyArgTyrArgLeuProGlyArgAlaCysArgArgValProAlaAlaLeuArgArgPro 250
QY 292 CGTAGTGGTGGTGGCTCCCGAGGTGAAGCCCTCGGCGAGGAGCTGCTGAAGT----- 239
Db 251 ProThrAlaTyr---ArgProArg-----ArgGlyGlyTyrCysTyrGlyAspGly 266
QY 238 -----AGTCGGGATGCCGACACGCTACTGTATGAAGCGCT 203
Db 267 ArgLeuArgArgIleArgGlyArgGlySerGlyCysAlaAla----- 280
QY 202 TCGAGCCGTACATCGCAGGAGTGGACAGATGTGAAGCGGAGGCGGCGCGC--- 146
Db 281 ---AlaArgSerCysArgTyrTyr-----CysGlyArgProArgProGlySerAlaAla 297
QY 145 -----CCTCGATCACCTCGATCTTCATCTCTCGGTGCTCGAGGGGGTTCCT 95
Db 298 GlnSerTyrProValSerSerGlnAlaSerGlyGluValSerProCysGlySerCysArg 317
QY 94 CGC-----CCTTGGCGGTGCACTTGAAGTAGTGGCTTACCGTGGCTCGATGG 44
Db 318 ArgSerArgAlaAlaCysGlnSerProLysArgSerLeuSerSerAlaSerProGly--- 336
QY 43 TGTCTCTGAAGGCGCATGCT 23
Db 337 -----GlyAlaTyrSer 340

RESULT 10
US-09-252-991A-26899
; Sequence 26899, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26899
; LENGTH: 247
; TYPE: PPT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26899

Alignment Scores:
Pred. No.: 4.92e-11 Length: 247

```


Score: 232.50 Matches: 88
Percent Similarity: 40.39% Conservative: 15
Best Local Similarity: 34.51% Mismatches: 79
Query Match: 17.88% Indels: 73
DB: 4 Gaps: 16

US-10-081-864-21 (1-707) x US-09-252-991A-26899 (1-247)

```
QY 8 CTTCCCTGTCAGCGAGCACCCTCTCAGGACCA-----CCATCGAGGGCA 55
Db 3 ProSerCysSerPro--ProSerProCysSerProLeuArgArgTrpProSerProThr 21
QY 56 CCGTGAACGGCCCTACTTCAAGTGCACCGCGAGGGCG----- 94
Db 22 ProAlaThrAlaAsnThrAlaTrpSerProAlaSerProIleArgSerProAlaSerThr 41
QY 95 -----AGGGCAACCCCTCGAGGGCACCCAGAGATGAAGATCGAGGTGATCGAGG 145
Db 42 ThrCysTrpArgCysSerProSer--AlaSerGlyProProSerAsnArgLupProArg 60
QY 146 GCGGCCCTGCTGCTGCTGCTTCCATCC---TGTCACCTCTCTGATGATCGGCTCCA 202
Db 61 ValSerProCysProAlaTrpSerSerAlaProCysTrpSerAlaGlyCysSerAlaSer 80
QY 203 AGGCTTTATCA-----AGTACGTGTCGGGCTATCCCGACTACTTCAAGCAGTCCCTCG 256
Db 81 ArgAlaCysAsnCysArgSerTrpIleProAlaSerProProArgCys-----Ser 97
QY 257 CCGAGGCTTCACTGGAGCGGACCCACCATCTACGAGGAGCGGCTTCTGACCGCCC 316
Db 98 ProSerAlaSerAlaTrpProSerProCysAlaArgCysArgTrpArgTrpProPro 117
QY 317 ACCAGGACCTCTCTGAGCGGCGACTGCTGCTGTACAGGTGAGATCCTGGGCAACA 376
Db 118 ProArgCysSerProTrp-----Pro 124
QY 377 ACTTCCCGCGAGCGGCGGCTGATGAGAACAGCGCGGCTGCTGG----- 424
Db 125 ThrGlySerProThrAlaTrpSerCysArgThrSerProAlaProGlyCysThrArgSer 144
QY 425 -----AGCCTTCACCGAGATCGT-----ACGAGGTGGAGCGGCTGCTGC 466
Db 145 ValSerSerProProProProArgCysThrProAlaThrArgTrp-----Cys 161
QY 467 GCGGCC-----AGTCCCTGATGGCCCTGGAGTGC 496
Db 162 AlaSerTrpArgArgProArgProTrpTrpTyrrValSerProGlyTrpProArgArgAla 181
QY 497 CCGGCGGTGCGCACCTGACCTGCACACCTACCGCTCAAGAGCCGCTC 556
Db 182 ProGly-----SerGlyCysTrpProAlaGluProProArgLeuPro--- 195
QY 557 CCGCCCTGAAGTCCCGGCTTCCACTTCGAGGACCCCGCATCGAGATCCTGGAGGAG 616
Db 196 -----GlnArgLeuIleSerThrArgArgProTrpCysThrArgArg---ArgPro 211
QY 617 TGG-----AGAAGGGCAAGTGTCTACAGCAGTACG 646
Db 212 TrpValCysAlaSerIleAlaArgAlaSerGlyArgSerSerThr 226
```

RESULT 11

US-09-472-065A-5
Sequence 5, Application US/09472065A
Patent No. 6638732

GENERAL INFORMATION:
APPLICANT: Evans, Krista

TITLE OF INVENTION: Mutants of Green Fluorescent Protein
FILE REFERENCE: 0942.402002

CURRENT FILING DATE: 1999-12-23
PRIORITY FILING DATE: 1997-11-14

PRIOR FILING DATE: 1997-11-14
CURRENT APPLICATION NUMBER: US 60/030,935

PRIOR FILING DATE: 1996-11-15

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 238

TYPE: PRT

ORGANISM: Aequorea victoria, A1 mutant

US-09-472-065A-5

Alignment Scores:

Pred. No.: 1,03e-10 Length: 238
Score: 228.50 Matches: 60
Percent Similarity: 47.47% Conservative: 43
Best Local Similarity: 27.65% Mismatches: 77
Query Match: 17.58% Indels: 38
DB: 4 Gaps: 7

US-10-081-864-21 (1-707) x US-09-472-065A-5 (1-238)

```
QY 13 CTGTCGACGAGACCATGCTTCAGGACCATCCCTCAGGAGCGACCGTGAACGGCCACTAC 72
Db 7 LeuPheThrGlyValValProIleLeuValGluLeuAspGlyAspValAsnGlyHisys 26
QY 73 TTCAGTGCACCGGACGAGCGGAGGCAACCCCTCGAGGCGACCCAGGAGATGAAGATC 132
Db 27 PheSerValSerGlyGluGlyGluGlyAspAlaThrTrpGlyLysLeuThrLeuLysPhe 46
QY 133 GAGGTGATCGAGGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 192
Db 47 ---IleCysThrThrGlyLysLeuProValProTrpProThrLeuValThrThrCysAla 65
QY 193 TACGGCTCAAGGCTTCATCAAGTACGTGCTCGGCTATCCCTGCTGCTGCTGCTGCTGCTG 246
Db 66 TyrGlyValGlnCysPheSerArgTrpProAspHisMetLysGlnHisAspPheLys 85
QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGCGCACCATCCACCTACGAGGAGCGGCTTC 306
Db 86 SerAlaMetProGluGlyTyrrValGlnGluArgThrIlePhePheLysAspGlyAsn 105
QY 307 CTGACCGCGCCACGAGCACCTCCCTGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 366
Db 106 TyrLysThrArgAlaGluValPheGluGlyAspThrLeuValAsnArgIleGluLeu 125
QY 367 CTGGGCAACAACTTCCCGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 411
Db 126 LysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysLeuGluTyrrAsnTyr 145
QY 412 -----GCCGCGCTGGGAGCCCTCCACCGAGATCGTGTAC 447
Db 146 AsnSerHisAsnValTyrrIleMetAlaAspLysGlnLysAsnGlyIleLysValAsnPhe 165
QY 448 GAGGTG-----GACGCGCTGCTGCGCGG----- 470
Db 166 LysIleArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrrGlnGlnAsn 185
QY 471 ---CGAGTCCCTGATGCGCTGAGTGCCCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 527
Db 186 ThrProIleGlyAspGlyProValLeu-----LeuProAspAsn 198
QY 528 CACCACTACCGCTCCAGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 575
Db 199 HisTyrrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArg 215
```

RESULT 12

US-09-417-197-121

Sequence 121, Application US/09417197

Patent No. 6518021

GENERAL INFORMATION:

APPLICANT: Ole Thastrup, et al.

TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An

FILE REFERENCE: 3759-0110P

CURRENT APPLICATION NUMBER: US/09/417,197

```

; CURRENT FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 121
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFp-lkappab-kinase fusion
US-09-417-197-121

Alignment Scores:
Pred. No.: 1,5e-10 Length: 997
Score: 228.50 Matches: 65
Percent Similarity: 40.98% Conservativeness: 44
Best Local Similarity: 24.44% Mismatches: 86
Query Match: 17.58% Indels: 72
DB: 4 Gaps: 7

US-10-081-864-21 (1-707) x US-09-417-197-121 (1-997)

QY 13 CTGCTGACGAGACCATGCTTCCAGGACACCATCGAGGCGACCGTGACGGCCACTAC 72
Db LeuThrGlyValValProIleLeuValGluLeuAspGlyValAsnGlyHisLys 27

QY 73 TTCAAGTGCACCGGCAAGGGGAGGCAACCCCTCGAGGCGACCCAGGAGATGAAGATC 132
Db PheSerValSerGlyGluGlyAspAlaThrTyrglyLysLeuThrLeuLysPhe 47

QY 133 GAGGTGATCGAGGGGCGCCCTCCCTTCCATCTGCTCCATCTGCTCCATCTGCTGATG 192
Db ---IleCysThrThrGlyLysLeuProValProThrLeuValThrLeuThr 66

QY 193 TACGCTCCAAAGGCTTCATCAAGTAGCTGTCGGGATCCCTCCCTGCTTCAAG 246
Db TyrglyValGlnCysPheSerArgTyrglyProAspHisMetLysGlnHisAspPheLys 86

QY 247 CAGTCCCTCCCGAGGGTTCACCTGGGAGGCGACACCATCGAGGAGCGGGCTTC 306
Db SerAlaMetProGluGlyTyrglyValGlnGluArgThrIlePhePheLysAspGlyAsn 106

QY 307 CTGACCGCCACCGAGGACCATCTGAGCGGCGACCTGCTGGGTACAAAGGTGAAGATC 366
Db TyrglyThrArgAlaGluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeu 126

QY 367 CTGGCAACAATCTCCCGCGCGAGCGCCCTGTGATGAGCAAGAC 411
Db LysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysLeuGluTyrglyAsn 146

QY 412 -----GCCGGCGCTGGGAGCGCTCCACCGAGATCGTGTAC 447
Db AsnSerHisAsnValTyrglyLeuAlaAspLysGlnLysAsnGlyIleLysValAsn 166

QY 448 GAGGTG-----GACGGCGTGTGCTGGCGG 470
Db LysIleArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrglyGlnAsn 186

QY 471 ---CCAGTCCCTGATGCGCCCTGGAGTCCCGGGGTGCCACCTGACCTGCCACCTGCA 527
Db ThrProIleGlyAspGlyProValLeu-----LeuProAspAsn 199

QY 528 CACCACTACCGCTCCAAAGAGCC----- 551
Db HisTyrglyLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMet 219

QY 551 ----- 551
Db ValLeuLeuGluPheValThrAlaAlaGlyIleThrLeuGlyMetAspGluLeuTyrgly 239

QY 552 -----CGCTCCCGCTGAGATCGCCGGCTTCCATCTTCGAGGACCAACCGCATCGA 602
Db SerGlyLeuArgSerArgAlaGlnAlaSerAsnSerThrMetGluArgProProGlyLeu 259

```

```

QY 603 CATCTCGAGGAGGTGGA 620
Db 260 ArgProGlyAlaGlyGly 265

RESULT 13
US-09-252-991A-19529
; Sequence 19529, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19529
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19529

Alignment Scores:
Pred. No.: 1.45e-10 Length: 302
Score: 227.00 Matches: 89
Percent Similarity: 38.32% Conservativeness: 16
Best Local Similarity: 32.48% Mismatches: 81
Query Match: 17.48% Indels: 88
DB: 4 Gaps: 17

US-10-081-864-21 (1-707) x US-09-252-991A-19529 (1-302)

QY 47 TCGAGGGCACCG---TGAACGGCCACTACTTCAAGT-----GCACCG 85
Db 1 SerLysAlaProGlyGluThrAlaLysThrArgSerProGlySerProArgThrAlaPro 20

QY 86 GCAAGGGGAGGGCAACCCCTCGAGGCGACCCAGGAGATGAAGATCGAGTGTGAGG 145
Db 21 ValAlaGlyArgAlaThrProAlaArgArgProAlaArg----- 33

QY 146 GCGGCGCCCTGCGCTTCGCCTCCACATCTGCTCCACCTCTGATGTACGCTCCAAAG 205
Db 34 -----ProCys-----ArgCysArgLysProAlaAlaThrThr----- 44

QY 206 CCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAGCAGTCCCTCCCGAGGCT 265
Db 45 ArgSerCysSerThrArgGlyAlaThrArgCysArgAlaSerTrpProSer---AlaCys 63

QY 266 TCACCTGGGAGGCGACCCACCATCGAGGCGGGCTTCTGACCGCCACCGAGGACA 325
Db 64 SerProAlaSerPheProProArgArg-----ProProAlaArgCys 78

QY 326 CCTCCTGACGCGGCACTGCTGTGT-----ACAAGGTGAAGATCTCGGCA 373
Db 79 ArgAlaThrAlaArgThrAlaThrCysArgArgAlaThrArgArgAlaAlaAla 98

QY 374 ACACCTCCCG----- 385
Db 99 ValSerProProMetAlaSerThrAlaArgAlaAlaGlyAlaSerProSerArgTrp 118

QY 386 CGAGCGGCGCGCTGTGTGAGAACAGCGCGCGCTGGAGCCCTCCACCGAGATCGTGT 445
Db 119 ProArgAlaSerArgCysSerProArgProThrSerSerProPro-----Cys 135

QY 446 ACAGGTGAGC-----GGTCTCGCGCGCGGCGGCTGTGTCGTCGTCGTCGTCG 487
Db 136 CysArgTrpProArgArgProProSerGlySerAlaSerMetProAlaThrTrpPro 155

QY 488 -----TGG-----AGTGGCGCGCGCTGCGCCACCTGA 514

```

```
Db 156 ArgProArgTirArgProAlaGlyAlaAsnArgSerSerAlaSerArgMetAlaAla 175
QY 515 CTGCGCCACCTGCACA-----CCACCTACCGCTCCCAAGAGCCCG----- 553
Db 176 AlaThrThrCysSerAlaSerArgProAlaAlaGlyMetValProArgAlaLeu 195
QY 554 ---CCTCGCGCTGAGAGCCCGGCTTCCACTTCGAGGACCAACCCATCGAGATCCTGG 610
Db 196 SerProProAnThrLysAlaAlaArgArgSerArgLysArgSerArgAlaProThr 215
QY 611 AGGAGTGGAGAGGCAAGT-----CCTACAAGCAGTACG 646
Db 216 ArgArgSerSerArgAlaSerThrThrSerAlaSerSerArgAlaThrThrAlaThr 235
QY 647 AGCGCGCGTGGCGCGCTACTGCGACGCGCGCCCTCCCAAGC 688
Db 236 ArgThrProThrAla-----ProProThrProSer 246

RESULT 14
US-09-417-197-39
; Sequence 39, Application US/09417197
; Patent No. 6518021
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP, et al.
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An I
; FILE REFERENCE: 3759-0110P
; CURRENT APPLICATION NUMBER: US/09/417,197
; CURRENT FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 39
; LENGTH: 631
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-Erfk1 fusion construct
US-09-417-197-39

Alignment Scores:
Pred. No.: 2,32e-10 Length: 631
Score: 225.50 Matches: 73
Percent Similarity: 41.28% Conservative: 50
Best Local Similarity: 24.50% Mismatches: 94
Query Match: 17.35% Indels: 82
DB: 4 Gaps: 11

US-10-081-864-21 (1-707) x US-09-417-197-39 (1-631)
QY 13 CTGCTGACCGAGACCATGCGCTTTCAGGACCAACATCGAGGCGCACCGTGAACGCGCACTAC 72
Db 8 LeuPheThrGlyValValProLeuLeuValGluLeuAspGlyAspValAsnGlyHisLys 27
QY 73 TTCAGTGCACCGCAAGCGGCGGCAACCCCTCGAGGCGCACCGAGATGAGATC 132
Db 28 PheSerValSerGlyGlyGluGlyAspAlaThrTyrGlyLysLeuThrLeuLysPhe 47
QY 133 GAGGTGATCGAGGCGCGCCCTTCGCTTCCACATCTCCACATCTCCACCTCTCGCATG 192
Db 48 ---IleCysThrThrGlyLysLeuProValProThrProThrLeuValThrThrLeuThr 66
QY 193 TAGCGTCCAGCGCTTCATCAAGTACGTGTCGGGATCCCG-----GACTACTTCAAG 246
Db 67 TyrGlyValGlnCysPheSerArgTyrProAspHisMetLysGlnHisAspPhePheLys 86
QY 247 CAGTCCCTCCCGAGCGCTTCACCTGGGAGCGGCAACCCACCTACGAGGACCGCGGCTTC 306
Db 87 SerAlaMetProGluGlyTyrValGlnGluArgThrIlePhePheLysAspGlyAsn 106
QY 307 CTGACCGCCCAACAGGACACCTCCCTGGACGCGACTCGCTGTGTACAGGTGAAGATC 366
Db 107 TyrLysThrArgAlaGluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeu 126
```

```
QY 367 CTGGGCAACAACACTTCCCGCGCGAGCGGCGCGCTGTGTGAGCAACAAG----- 411
Db 127 LysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysLeuGluTyrAsnTyr 146
QY 412 -----GCGGCGCGCTGGGAGCCCTCCACGAGATCGTGTAC 447
Db 147 AsnSerHisAsnValTyrIleMetAlaAspLysGlnLysAsnGlyIleLysValAsnPhe 166
QY 448 GAGGTG-----GACGCGGTGCTGCGCG----- 470
Db 167 LysIleArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsn 186
QY 471 ---CCAGTCCCTGATGCGCTGAGTCCCGCGGCTCGCCACCTGACCTGCCACCTGCA 527
Db 187 ThrProIleGlyAspGlyProValLeu-----LeuProAspAsn 199
QY 528 CACCACCTACCGCTCCCAAGAGCCCGCTCGC---CCTGAAGATGCGCG----- 575
Db 200 HisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMet 219
QY 576 -----CTTCCACTT 584
Db 220 ValLeuLeuGluPheValThrAlaAlaGlyIleThrLeuGlyMetAspGluLeuTyrLys 239
QY 585 CGAGGACCAACCGCATCGA-----CAT 605
Db 240 SerGlyLeuArgSerArgAlaGlnAlaSerAsnSerThrMetAlaAlaAlaAlaGln 259
QY 606 CTGAGGAGGTGAGAGGCAAGTGTACAGCAGTACGAGCGCGCGCTGGGCGCG--- 662
Db 260 GlyGlyGlyGlyGlyGluProArgThrGluGlyVal-----GlyProGly 275
QY 663 -----CTACTCGAGCGCGCGCGCTCCCAAGCTGGGCCACAA 698
Db 276 ValProGlyGluValGluMetValLysGlyGlnProPheAspValGlyProArg 293

RESULT 15
US-09-252-991A-31608
; Sequence 31608, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31608
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31608

Alignment Scores:
Pred. No.: 2,13e-10 Length: 322
Score: 225.00 Matches: 89
Percent Similarity: 38.05% Conservative: 24
Best Local Similarity: 29.97% Mismatches: 96
Query Match: 17.31% Indels: 88
DB: 4 Gaps: 15

US-10-081-864-21 (1-707) x US-09-252-991A-31608 (1-322)
QY 5 CCGCTCCCTCTGCTGACCGAGA-----CCA 28
Db 15 ProThrThrCysSerProArgAsnArgLysAsnSerProThrSerArgThrGlyPro 34
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:36:21 ; Search time 44.913 Seconds
(without alignments)
9873.883 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 1300

Sequence: 1 ggaatcgctccctgctgac.....ctgggccacaactgaagctt 707

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 2582470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10081864/runat_29072004_150508_25599/app_query.fasta_1.1742
-DB=Published Applications AA -QMT=fastan -SUFFIX=rapp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -LOCAL=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USPR=US10081864 @CGN 1 1 21 @runat_29072004_150508_25599
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -55599
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description

RESULT 1

US-10-006-922-40
; Sequence 40, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tskikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; FILE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556

ALIGNMENTS

Handwritten signature and initials.

1	1257	96.7	231	13	US-10-006-922-40	Sequence 40, Appl
2	1257	96.7	231	14	US-10-081-864-22	Sequence 22, Appl
3	1251	96.2	232	14	US-10-081-864-24	Sequence 24, Appl
4	1226	94.3	232	13	US-10-006-922-42	Sequence 42, Appl
5	1219	93.8	232	14	US-10-155-809-2	Sequence 2, Appl
6	1219	93.8	232	15	US-10-370-570-59	Sequence 59, Appl
7	1218	93.7	232	14	US-10-155-809-4	Sequence 4, Appl
8	1216	93.5	232	13	US-10-006-922-14	Sequence 14, Appl
9	1216	93.5	232	14	US-10-081-864-10	Sequence 10, Appl
10	1213	93.3	232	16	US-10-724-178-18	Sequence 18, Appl
11	1206	92.8	232	14	US-10-155-809-6	Sequence 6, Appl
12	1045	80.4	197	16	US-10-724-178-1047	Sequence 1047, Ap
13	961	73.9	185	16	US-10-724-178-1065	Sequence 1065, Ap
14	868	66.8	168	16	US-10-724-178-1061	Sequence 1061, Ap
15	785	60.4	152	16	US-10-724-178-1053	Sequence 1053, Ap
16	780.5	60.0	127	14	US-10-155-809-16	Sequence 16, Appl
17	770.5	59.3	226	9	US-09-976-673-10	Sequence 10, Appl
18	770.5	59.3	227	14	US-10-155-809-14	Sequence 14, Appl
19	769.5	59.2	226	9	US-09-976-673-8	Sequence 8, Appl
20	769.5	59.2	226	9	US-09-976-673-16	Sequence 16, Appl
21	769.5	59.2	226	9	US-09-976-673-18	Sequence 18, Appl
22	768.5	59.1	227	9	US-09-976-673-24	Sequence 24, Appl
23	767.5	59.0	227	9	US-09-976-673-6	Sequence 6, Appl
24	766.5	59.0	226	9	US-09-976-673-26	Sequence 26, Appl
25	765.5	58.9	227	9	US-09-976-673-12	Sequence 12, Appl
26	762.5	58.7	227	9	US-09-976-673-2	Sequence 2, Appl
27	762.5	58.7	227	9	US-09-976-673-14	Sequence 14, Appl
28	757.5	58.3	227	14	US-10-155-809-10	Sequence 10, Appl
29	756.5	58.2	227	14	US-10-155-809-12	Sequence 12, Appl
30	753.5	58.0	227	9	US-09-976-673-4	Sequence 4, Appl
31	742	57.1	215	14	US-10-155-809-8	Sequence 8, Appl
32	718	55.2	135	16	US-10-724-178-1051	Sequence 1051, Ap
33	707	54.4	148	15	US-10-370-570-57	Sequence 57, Appl
34	643	49.5	121	16	US-10-724-178-1059	Sequence 1059, Ap
35	637	49.0	228	15	US-10-370-570-60	Sequence 60, Appl
36	570	43.8	111	16	US-10-724-178-1057	Sequence 1057, Ap
37	567.5	43.7	225	16	US-10-423-688A-41	Sequence 41, Appl
38	553	42.5	225	14	US-10-244-779-2	Sequence 2, Appl
39	551.5	42.4	225	15	US-10-442-148A-7	Sequence 7, Appl
40	551.5	42.4	227	15	US-10-442-148A-7	Sequence 7, Appl
41	551.5	42.4	236	15	US-10-370-570-62	Sequence 62, Appl
42	551.5	42.4	239	15	US-10-314-936-4	Sequence 4, Appl
43	550.5	42.3	225	15	US-10-442-148A-8	Sequence 8, Appl
44	550.5	42.3	225	16	US-10-421-258-8	Sequence 8, Appl
45	550.5	42.3	226	14	US-10-724-178-16	Sequence 16, Appl
			226	14	US-10-121-258-6	Sequence 6, Appl

```

; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Anemonia sulcata
US-10-006-922-40

Alignment Scores:
Pred. No.: 1.41e-76 Length: 231
Score: 1257.00 Matches: 231
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.69% Indels: 0
DB: 13 Gaps: 0

US-10-081-864-21 (1-707) x US-10-006-922-40 (1-231)
QY 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGCACCGTGAACGGC 66
DB 1 AlaserLeuLeuThrGluThrMetProPheArgThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCGAGGCAACCCCTTCGAGGCGCACCCAGGAGATG 126
DB 21 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCCCTGTCACCTCC 186
DB 41 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCGGGCATCCCGACTACTTCAAG 246
DB 61 CysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGCTTCACTGGGAGCGCACCAACCCCTTCGAGGCGCACCCAGGAGATG 306
DB 81 GlnSerLeuProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACACGAGCACCTCCCTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 366
DB 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
QY 367 CTGGGCAACAACCTTCCCGCGAGTGTACGAGTGTGAGCGGCGGCGGCGGCGGCGGCGGCTTC 426
DB 141 ProSerThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
QY 487 CTGAGTGCCTCCCGCGGCTCCACCTGCACCTGCACACCAACCTACCGCTCCAAAG 546
DB 161 LeuGluCysProGlyGlyArgHisLeuThrCysHisLeuHisIleThrThrTyrArgSerLys 180
QY 547 AAGCCCGCTCCCGCTTCAGATGCGCGGCTTCACCTTCGAGGACCAACCGCATCGAGATC 606
DB 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 200
QY 607 CTGAGAGGTGGAGAGGCGCAAGTGTCTACAAGCAGTACGAGGCGGCGGCGGCGGCGGCTAC 666
DB 201 LeuGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyAspTyr 220
QY 667 TGGACCGCGCGCTCCAAAGCTGGGCGCAAC 699
DB 221 CysAspAlaAlaProSerLysLeuGlyHisAsn 231

```

RESULT 2

```

US-10-081-864-22
; Sequence 22, Application US/10081864
; Publication No. US20030022287A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: non-aggregating mutant
US-10-081-864-22

```

```

Alignment Scores:
Pred. No.: 1.41e-76 Length: 231
Score: 1257.00 Matches: 231
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.69% Indels: 0
DB: 14 Gaps: 0

```

```

US-10-081-864-21 (1-707) x US-10-081-864-22 (1-231)
QY 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGCACCGTGAACGGC 66
DB 1 AlaserLeuLeuThrGluThrMetProPheArgThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCGAGGCAACCCCTTCGAGGCGCACCCAGGAGATG 126
DB 21 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCCCTGTCACCTCC 186
DB 41 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGGCTCCAAAGGCTTCATCAAGTACGTGTCGGGCATCCCGACTACTTCAAG 246
DB 61 CysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGCTTCACTGGGAGCGCACCAACCCCTTCGAGGCGCACCCAGGAGATG 306
DB 81 GlnSerLeuProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACACGAGCACCTCCCTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 366
DB 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
QY 367 CTGGGCAACAACCTTCCCGCGAGTGTACGAGTGTGAGCGGCGGCGGCGGCGGCGGCTTC 426
DB 121 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 140
QY 427 CCCTCCACCGAGTGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 486
DB 141 ProSerThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
QY 487 CTGAGTGCCTCCCGCGGCTCCACCTGCACCTGCACACCAACCTACCGCTCCAAAG 546
DB 161 LeuGluCysProGlyGlyArgHisLeuThrCysHisLeuHisIleThrThrTyrArgSerLys 180
QY 547 AAGCCCGCTCCCGCTTCAGATGCGCGGCTTCACCTTCGAGGACCAACCGCATCGAGATC 606

```

Db 181 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluLe 200
QY 607 CTGGAGGAGTGGAGAGCGGAGTGTACAGAGCAGTACGAGCGCGCGTGGCGCGCTAC 666
Db 201 LeuGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaAlaValGlyArgTyr 220
QY 667 TGGCAGCGCGCGCGCTCCCAAGCTGGCGCACAC 699
Db 221 CysAspAlaAlaProSerLysLeuGlyHisasn 231
RESULT 3
US-10-081-864-24
; Sequence 24, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey
; APPLICANT: Lukanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexander
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: non-aggregating mutant
US-10-081-864-24
Alignment Scores:
Pred. No.: 232
Score: 1251.00
Length: 3.57e-76
Matches: 230
Conservative: 0
Percent Similarity: 99.57%
Best Local Similarity: 99.57%
Mismatch: 1
Indels: 0
Query Match: 14
Gaps: 0
US-10-081-864-21 (1-707) x US-10-081-864-24 (1-232)
QY 7 GCCTCCTGCTGACCGAGACCATGCGCTTCAGGACACCATCGAGGCGCCGCTGACGCGC 66
Db 2 AlaSerLeuLeuThrGluThrMetProPheArgThrThrIleGluGlyThrValasnGly 21
QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTCGAGGCGCACCGGAGATG 126
Db 22 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProLeuGluGlyThrGlnGluMet 41
QY 127 ARGATCGAGTGTACGAGGCGCGCGCTGCTGCTTCACATCTGCTCCACTGCTCCACTCC 186
Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGTACGCGCTCCAGGCGCTTCATCAGTACGTGTGCGGATCCCGGACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysAlaPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCGCTTCACTGCGAGCGGCGCACACCATCGAGGCGCGCGCTTC 306
Db 82 GlnSerLeuProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCGCGCCACAGGACACCTCCCTCGAGCGGCGACTGCTGCTGTGTACAGGTGAAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121

QY 367 CTGGGCAACAACCTTCCCGCGCGAGCGCGCTGATGCAGAACAGGCGCGCGCTGGGAG 426
Db 122 LeuGlyAsnAspPheProAlaAspGlyProValMetGlnasnLysAlaGlyArgTrpGlu 141
QY 427 CCCTCCACCGAGATCGTGTACGAGGTGGACGGCGTCTCGCGGCGCAGTCCCTGTATGGCC 486
Db 142 ProSerThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerSerMetAla 161
QY 487 CTGGAGTGGCGCGCGTGGCGCGCGCTGACCTGCCACCTGCACACCATCGCTCCCAAG 546
Db 162 LeuGluCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCCCGCTCCCGCTGAGATGCGCGCGCTTCCACTTCCAGGACCGCGCATCGAGATC 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluLe 201
QY 607 CTGGAGGAGTGGAGAGCGGCAAGTGTCTACAGCAGTACGAGCGCGCGCTGAGATC 666
Db 202 LeuGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaAlaValGlyArgTyr 221
QY 667 TGGCAGCGCGCGCGCTCCCAAGCTGGCGCACAC 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisasn 232
RESULT 4
US-10-006-922-42
; Sequence 42, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Anemonia sulcata
US-10-006-922-42
Alignment Scores:
Pred. No.: 1.73e-74
Score: 1226.00
Length: 232
Matches: 223
Conservative: 4
Percent Similarity: 98.27%
Best Local Similarity: 98.54%
Mismatch: 4
Query Match: 94.31%
Indels: 0
Gaps: 0
US-10-081-864-21 (1-707) x US-10-006-922-42 (1-232)
QY 7 GCCTCCTGCTGACCGAGACCATGCGCTTCAGGACACCATCGAGGCGCGCGCTGACGCGC 66
Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValasnGly 21

QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTCGAGGGCACCCGAGGATG 126
 Db 22 HistyrPheLysCysThrGlyValGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCGAGGTGATCGAGGGCGGGCCCTGCTCCCTTCCACATCTCTGTCACCTCC 186
 Db 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
 QY 187 TGCATGTACGGCTCCACAGGCTTCATCAAGTACGTGTCCGCTCCCGGACTACTTCAAG 246
 Db 62 CysMetTyrGlySerLysAlaPheileLysTyrValSerGlyileProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGCGGACACCACTACGAGGACGGCGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACAGGACCTCCCTCGAGCGGACCACTACGAGGACGGCGCTTC 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 367 CTGGGCAACACTTCCCGCGGCGGCTGATGAGCAAGCAAGCGCGCGCTGGAG 426
 Db 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTCCACCGAGTGTGTAGAGGTGAGCGGTGCTGCGCGGCGGCTGATGAGTGGCC 486
 Db 142 ProSerThrGluileValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGAGGTGCCCGCGCTGCGCCACTGACCTGACCTGACCTGACCTGACCTGACCTGAC 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCCCGCTCCCGCTGGAAGTGCCTCAAGTGGGCGCAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

No CD

RESULT 5
 US-10-155-809-2
 ; Sequence 2, Application US/10155809
 ; Publication No. US20030092884A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lukyanov, Sergey
 ; APPLICANT: Lukyanov, Konstantin
 ; APPLICANT: Chudakov, Dmitry
 ; TITLE OF INVENTION: Kindling Fluorescent Proteins and
 ; FILE REFERENCE: CLON-073
 ; CURRENT APPLICATION NUMBER: US/10/155,809
 ; CURRENT FILING DATE: 2002-05-24
 ; PRIOR APPLICATION NUMBER: 60/293,752
 ; PRIOR FILING DATE: 2001-05-25
 ; PRIOR APPLICATION NUMBER: 60/329,176
 ; PRIOR FILING DATE: 2001-10-11
 ; PRIOR APPLICATION NUMBER: 10/006,922
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 09/210,330
 ; PRIOR FILING DATE: 1998-12-11
 ; PRIOR APPLICATION NUMBER: 09/976,673
 ; PRIOR FILING DATE: 2001-10-12
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 232
 ; TYPE: PRT
 ; ORGANISM: anthozoa

US-10-155-809-2

Alignment Scores:

Pred. No.: 232
 Score: 1219.00
 Percent Similarity: 97.84%
 Best Local Similarity: 95.67%
 Query Match: 93.77%
 Indels: 0
 Gaps: 0

US-10-081-864-21 (1-707) x US-10-155-809-2 (1-232)

QY 7 GCCTCCCTCGTACGACGACCATGCTTCAGGACCACTCGAGGGCACCGTCAAGGCG 66
 Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrileGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTCGAGGGCACCCGAGGATG 126
 Db 22 HistyrPheLysCysThrGlyValGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCGAGGTGATCGAGGGCGGGCCCTGCTCCCTTCCACATCTCTGTCACCTCC 186
 Db 42 LysileGluValileGluGlyGlyProLeuProPheAlaPheHisileLeuSerThrSer 61
 QY 187 TGCATGTACGGCTCCACAGGCTTCATCAAGTACGTGTCCGCTCCCGGACTACTTCAAG 246
 Db 62 CysMetTyrGlySerLysThrPheileLysTyrValSerGlyileProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGGCTTCACCTGGAGCGGACCACTACGAGGACGGCGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACCGCCACAGGACCACTCCCTCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 366
 Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysile 121
 QY 367 CTGGGCAACACTTCCCGCGGCGGCTGATGAGCAAGCAAGCGCGCGCTGGAG 426
 Db 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTCCACCGAGTGTGTAGAGGTGAGCGGTGCTGCGCGGCGGCTGATGAGTGGCC 486
 Db 142 ProAlaThrGluileValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGAGGTGCCCGCGCTGCGCCACTGACCTGACCTGACCTGACCTGACCTGACCTGAC 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCCCGCTCCCGCTGGAAGTGCCTCAAGTGGGCGCAAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 6

US-10-370-570-59
 ; Sequence 59, Application US/10370570
 ; Publication No. US20030219717A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DAHL, Soren Weis et al.
 ; TITLE OF INVENTION: FLUOROPHORE COMPLEMENTATION PRODUCTS
 ; FILE REFERENCE: 3759-0130P
 ; CURRENT APPLICATION NUMBER: US/10/370,570
 ; CURRENT FILING DATE: 2003-02-24
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 59

; LENGTH: 232
; TYPE: PRT
; ORGANISM: Anemonia sulcata
US-10-370-570-59

Alignment Scores:

Pred. No.: 5,11e-74 Length: 232
Score: 1219.00 Matches: 221
Percent Similarity: 97.84% Conservative: 5
Best Local Similarity: 95.67% Mismatches: 5
Query Match: 93.77% Indels: 0
DB: 15 Gaps: 0

US-10-081-864-21 (1-707) x US-10-370-570-59 (1-232)

```
QY 7 GCCTCCCTGTGACCGAGACCATGCGCTTCAGGACCAACCATCGAGGCGCCCTGTAACGCG 66
Db 2 AlaserPheLeuLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
QY 67 CACTACTTCAGTCCAGCGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGG 126
Db 22 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCGAGGTGATCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 186
Db 42 LysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGTACGCTCCAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyLeuProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCGCTTCACCTGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 306
Db 82 GlnSerPheProGluGlyPheThrTyrGluArgThrThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCGCCCGACGAGACACTCCCTCGAGCGGCGAGTGTCTGTGTACAGGTGAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
QY 367 CTGGGCAACAACTTCCCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGG 426
Db 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTyrGlu 141
QY 427 CCCTCCACCGAGATCGTGTACGAGGTGAGCGGTGTCTCGCGCGCGAGTCCCTGTGATGCC 486
Db 142 ProAlaThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCCCGCTCCCGCTCGAGATGCCCGGCTTCCACTTCAGGACCAACCGCATCAGATC 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
QY 607 CTGAGGAGGTGGAGAGGCGCAAGTGTCTACAGCAGTACGAGCGCGCGCGCGCGCGCTAC 666
Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
QY 667 TCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232
```

RESULT 7

; Sequence 4, Application US/10155809
; Publication No. US20030092884A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Chudakov, Dmitry
; TITLE OF INVENTION: Kindling Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Their Use

; FILE REFERENCE: CLON-073
; CURRENT APPLICATION NUMBER: US/10/155,809
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/293,752
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/329,176
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/210,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/976,673
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 232
; TYPE: PRT
; ORGANISM: anthozoa
US-10-155-809-4

Alignment Scores:

Pred. No.: 5,97e-74 Length: 232
Score: 1218.00 Matches: 221
Percent Similarity: 97.40% Conservative: 4
Best Local Similarity: 95.67% Mismatches: 6
Query Match: 93.69% Indels: 0
DB: 14 Gaps: 0

US-10-081-864-21 (1-707) x US-10-155-809-4 (1-232)

```
QY 7 GCCTCCCTGTGACCGAGACCATGCGCTTCAGGACCAACCATCGAGGCGCCCTGTAACGCG 66
Db 2 AlaserPheLeuLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
QY 67 CACTACTTCAGTCCAGCGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGG 126
Db 22 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCGAGGTGATCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 186
Db 42 LysIleGluValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGTACGCTCCAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyLeuProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCGCTTCACCTGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 306
Db 82 GlnSerPheProGluGlyPheThrTyrGluArgThrThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCGCCCGACGAGACACTCCCTCGAGCGGCGAGTGTCTGTGTACAGGTGAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
QY 367 CTGGGCAACAACTTCCCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGG 426
Db 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTyrGlu 141
QY 427 CCCTCCACCGAGATCGTGTACGAGGTGAGCGGTGTCTCGCGCGCGAGTCCCTGTGATGCC 486
Db 142 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCCCGCTCCCGCTCGAGATGCCCGGCTTCCACTTCAGGACCAACCGCATCAGATC 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
QY 607 CTGAGGAGGTGGAGAGGCGCAAGTGTCTACAGCAGTACGAGCGCGCGCGCGCGCGCTAC 666
Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
```

Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaAlaValGlyArgTyr 221
 QY 667 TCGGACGGCGCCCTCCCAAGCTGGCGGCACAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisasn 232

RESULT 8
 US-10-006-922-14
 ; Sequence 14, Application US/10006922
 ; Publication No. US20020197676A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lukyanov, Sergey A
 ; APPLICANT: Fradkov, Arcady F.
 ; APPLICANT: Labas, Yulii A.
 ; APPLICANT: Matz, Mikhail V.
 ; APPLICANT: Tersikh, Alexey
 ; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
 ; FILE REFERENCE: CLON-035CIP
 ; CURRENT APPLICATION NUMBER: US/10/006,922
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 09/120,330
 ; PRIOR FILING DATE: 1998-12-11
 ; PRIOR APPLICATION NUMBER: 09/457,898
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/458,144
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/458,477
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/457,556
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/444,338
 ; PRIOR FILING DATE: 1999-11-15
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14
 ; LENGTH: 232
 ; TYPE: PRT
 ; ORGANISM: Anemonia sulcata
 US-10-006-922-14

Alignment Scores:
 Pred. No.: 8,14e-74 Length: 232
 Score: 1216.00 Matches: 220
 Percent Similarity: 97.84% Conservative: 6
 Best Local Similarity: 95.24% Mismatches: 5
 Query Match: 93.54% Indels: 0
 DB: 13 Gaps: 0

US-10-081-864-21 (1-707) x US-10-006-922-14 (1-232)

QY 7 GCCTCCTGCTGACGAGACCATGCTTCAGGACACCATCGAGGACCATCGTGAACGGC 66
 Db 2 AlaserPheLeuLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTGCACCGGCGAGGCAACCCCTCGAGGACGAGGACGAGATG 126
 Db 22 HistyPheLysCysThrGlyGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCGAGTGATCGAGGCGGCGCCCTCGCTTCACATCTGTTCACCTCC 186
 Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
 QY 187 TGCATGTACGCTTCAAGGCTTCATCAAGTACGCTCGGATCCGATCCCGATCTCAAG 246
 Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGGCTTACCTGGGAGGCGACACCATCTACGAGGCGGGGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACGCGCCACAGACACTCCCTCGAGCGGAGCTGCTGGTGTTACAGGTGAAGATC 366

Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
 QY 367 CTGGGCAACAACCTTCCCCCGCCGACGCGCTGATGACAGAAACAGCGCGCGCTGGGAG 426
 Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
 QY 427 CCTCTCCACGAGATCGTGTACGAGTGGACGCGCTGCTGCGCGCCACTCCCTGATGCC 486
 Db 142 ProAlaThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
 QY 487 CTGAGTGCCTCCGCGCTCGCCACCTGACCTGCCACCTGACACCATCTACCGTCCCAAG 546
 Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
 QY 547 AAGCCCGCTCCGCTCGAAGATGCCCGCTTCCACTTCGAGGACCCCGCATCGAGATC 606
 Db 182 LysProAlaAlaAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
 QY 607 CTGAGGAGGTGGAGAGGCAAGTGTACAGCAGTACGAGGCGCGCTGGCGGCTAC 666
 Db 202 MetGluGluValGlyLysCysTyrLysGlnTyrGluAlaAlaValGlyArgTyr 221
 QY 667 TCGACGCGCGCCCTCCCAAGCTGGCGGCACAC 699
 Db 222 CysAspAlaAlaProSerLysLeuGlyHisasn 232

RESULT 9
 US-10-081-864-10
 ; Sequence 10, Application US/10081864
 ; Publication No. US20030022287A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lukyanov, Sergey
 ; APPLICANT: Yarushevich, Yuriy
 ; APPLICANT: Savitsky, Alexandr
 ; APPLICANT: Fradkov, Arcady
 ; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
 ; FILE REFERENCE: CLON-067
 ; CURRENT APPLICATION NUMBER: US/10/081,864
 ; CURRENT FILING DATE: 2002-06-19
 ; PRIOR APPLICATION NUMBER: 10/006,922
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 60/270,983
 ; PRIOR FILING DATE: 2001-02-21
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10
 ; LENGTH: 232
 ; TYPE: PRT
 ; ORGANISM: Anemonia sulcata
 US-10-081-864-10

Alignment Scores:
 Pred. No.: 8,14e-74 Length: 232
 Score: 1216.00 Matches: 220
 Percent Similarity: 97.84% Conservative: 6
 Best Local Similarity: 95.24% Mismatches: 5
 Query Match: 93.54% Indels: 0
 DB: 14 Gaps: 0

US-10-081-864-21 (1-707) x US-10-081-864-10 (1-232)

QY 7 GCCTCCTGCTGACGAGACCATGCTTCAGGACACCATCGAGGACCATCGTGAACGGC 66
 Db 2 AlaserPheLeuLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
 QY 67 CACTACTTCAAGTGCACCGGCGAGGCAACCCCTCGAGGACGAGGACGAGATG 126
 Db 22 HistyPheLysCysThrGlyGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41
 QY 127 AAGATCGAGTGATCGAGGCGGCGCCCTCGCTTCACATCTGTTCACCTCC 186
 Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
 QY 187 TGCATGTACGCTTCAAGGCTTCATCAAGTACGCTCGGATCCGATCCCGATCTCAAG 246
 Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
 QY 247 CAGTCCCTCCCGAGGGCTTACCTGGGAGGCGACACCATCTACGAGGCGGGGCTTC 306
 Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrThrTyrGluAspGlyGlyPhe 101
 QY 307 CTGACGCGCCACAGACACTCCCTCGAGCGGAGCTGCTGGTGTTACAGGTGAAGATC 366

Db 42 LysleuValIleGluGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyLeuProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGGCTTCACTGGAGCGGACCAACCACTACAGAGACGGCGCTTC 306
Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCGCCACCAAGGACACCTCCCTGGACGGGACTGCTGTGTACAAAGGTGAAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
QY 367 CTGGGCAACAATCTCCCGCGGACGGCCCGCTGTATGATGACAGAACAGCGCGCGCTGGAG 426
Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
QY 427 CCCCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGTGCGCGGCGGCTCCCTGTAGTGGC 486
Db 142 ProAlaThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGAGATGCGCGCGGCTGCGACCTGACCTGACACCTGACACCACTACCGCTCCAG 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCGCGCTCCCGCTCAAGATCCCGGCTTCCACTTCCAGTCCAGACCACTGAGATC 606
Db 182 LysProAlaAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
QY 607 CTGAGAGGTGAGAGGCAAGTGTACAAGCAGTACAGCGCGCGCTGGCGCGCTAC 666
Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
QY 667 TGCAGCGCGCGCTCCAGAGTGGGCGCACAAAC 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 10

US-10-724-178-18
; Sequence 18, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odysse Thera, Inc.
; APPLICANT: Micknick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; TITLE OF INVENTION: COMPLEMENTATION ASSAYS
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 232
; TYPE: PPT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: "KFP1" a mutated variant of kindling fluorescent protein asCP fro
; OTHER INFORMATION: m A. sulcata
US-10-724-178-18

Alignment Scores:

Pred. No.:	1,38-73	Length:	232
Score:	1213.00	Matches:	221
Percent Similarity:	96.97%	Conservative:	3
Best Local Similarity:	95.67%	Mismatches:	7
Query Match:	93.31%	Indels:	0
DB:	16	Gaps:	0

US-10-081-864-21 (1-707) x US-10-724-178-18 (1-232)
QY 7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCAACCGGTGAACGCG 66
Db 2 AlaSerLeuLeuThrGluThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGGCAACCCGAGGAGATG 126
Db 22 HisCysPheLysCysIleGlyLysGlyGlyAsnProPheGluGlyThrGlnGluMet 41
QY 127 AAGATCCAGGTGATCGAGGCGCGCCCTTCGCTTCCGCTTCCACATCTCTGTCCACCTCC 186
Db 42 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGACGGCTCCAAAGGCTTCATCAGTAGCTGTCCGGCATCCCGACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyLeuProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCTTCAGTGGAGCGGACCACTGAGGACCGCGCGCTTC 306
Db 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCGCCACCAAGGACACCTCCCTGGAGCGGACTGCTGTGTACAAAGGTGAAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
QY 367 CTGGGCAACAATCTCCCGCGGACGGCCCGCTGTATGATGACAGAACAGCGCGCGCTGGAG 426
Db 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysValGlyArgTrpGlu 141
QY 427 CCCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGTGCGCGCGCGCTCCCTGTATGGCC 486
Db 142 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGGAGTGGCGCGGCTGCGACCTGACCTGCGCACCTGCGCACCACTACCGCTCCAG 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCGCGCTCCCGCTCAAGATGCCCGCTTCCACTTCCAGTCCAGACCACTGAGATC 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspHisArgIleGluIle 201
QY 607 CTGGAGGAGGTGGAGAGGCAAGTGTACAAGCAGTACAGCGCGCGCTGGCGCGCTAC 666
Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
QY 667 TGCAGCGCGCGCTCCAGAGTGGGCGCACAAAC 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 11

US-10-155-809-6
; Sequence 6, Application US/10155809
; Publication No. US20030092884A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Chudakov, Dmitry
; APPLICANT: Lukyanov, Konstantin
; TITLE OF INVENTION: Kindling Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Their Use
; FILE REFERENCE: CLON-073
; CURRENT APPLICATION NUMBER: US/10/155,809
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/293,752
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/329,176
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/210,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/976,673
; PRIOR FILING DATE: 2001-10-12

```

; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 232
; TYPE: PRT
; ORGANISM: anthozoa
US-10-155-809-6

Alignment Scores:
Pred. No.: 3,846-73 Length: 232
Score: 1206.00 Matches: 219
Percent Similarity: 96.97% Conservativity: 5
Best Local Similarity: 94.81% Mismatches: 7
Query Match: 92.77% Indels: 0
DB: 14 Gaps: 0

US-10-081-864-21 (1-707) x US-10-155-809-6 (1-232)
QY 7 GCCTCCCTGTCGACCGACCATCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 66
Db 2 AlaSerPheLeuLysLysThrMetProPheLysThrThrIleGluGlyThrValAsnGly 21
QY 67 CACTACTTCAAGTCACCGGACGGGACGGGACCGGACCGGACCGGACCGGACGATG 126
Db 22 HisTyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGluGluMet 41
QY 127 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCTCCCTTCGACATCCCTGTCCACCTCC 186
Db 42 LysIleGluValIleGluGlyGlyProLeuProPhePheAlaPheHisIleLeuSerThrSer 61
QY 187 TGCATGTACGGCTCAAGGCTTCATCAAGTACGTGTCGGGATCCCGACTACTTCAAG 246
Db 62 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 81
QY 247 CAGTCCCTCCCGAGGCTTCACCTGGGAGCGCACCAACCATCCAGGACGGCGGCTTC 306
Db 82 GlnSerPheProGluGlyLeuThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 101
QY 307 CTGACCGCCACACGAGACACTCCCTGGACGGCGACTGCTGTGTACAGGTGAGATC 366
Db 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121
QY 367 CTGGGCAACAACCTCCCGCGGACGGCCCGTGTATCAGACAGGACCGGCGGCTGGAG 426
Db 122 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTrpGlu 141
QY 427 CCTCCACCGAGATCGTGTACGAGGTGGACGGGCTGCTGGCGGCGGCGGCTGATGGCC 486
Db 142 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161
QY 487 CTGGAGTCCCGCGGCTGCCACCTGATCGACCTGACACACCACTACCGTCCCAAG 546
Db 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrTyrArgSerLys 181
QY 547 AAGCCCGCTCCCGCTCGAAGATGCCCGCTTCCACTTCAGAGACCAACCGCATCGAGTC 606
Db 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAspTyrArgIleGlu 201
QY 607 CTGGAGAGGTGAGAAAGGCGAAGTGTCTACAGAGTACAGGCGCGCGCTGCGCTAC 666
Db 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221
QY 667 TGGAGCGCGCGCGCTCCCAAGCTGGGCCACAAC 699
Db 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 12
US-10-724-178-1047
; Sequence 1047, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen

; APPLICANT: MacDonald, Marne
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; TITLE OF INVENTION: COMPLEMENTATION ASSAYS
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1047
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KFP F2A, with Met added @ position 1
US-10-724-178-1047

Alignment Scores:
Pred. No.: 2,666-62 Length: 197
Score: 1045.00 Matches: 190
Percent Similarity: 97.96% Conservativity: 2
Best Local Similarity: 96.94% Mismatches: 4
Query Match: 80.38% Indels: 0
DB: 16 Gaps: 0

US-10-081-864-21 (1-707) x US-10-724-178-1047 (1-197)
QY 112 GGCACCCAGGAGATCAAGATCGAGGTGATCGAGGCGGCGCCCTGCCCTTCGCTCCAC 171
Db 2 GlyThrGlnGluMetLysIleGluValIleGluGlyGlyProLeuProPheAlaPheHis 21
QY 172 ATCTGTCCACTCTCTGATGTACGGCTCCAGGCTTCCTCAAGTACGTGTCGGGATC 231
Db 22 IleLeuSerThrSerCysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIle 41
QY 232 CCCGACTACTTCAAGCAGTCCCTCCCGAGGCTTCACCTGGGAGCGGCGGCGGCGGCTAC 291
Db 42 ProAspTyrPheLysGlnSerPheProGluGlyPheThrTrpGluArgThrThrTyr 61
QY 292 GAGGACGGCGCTTCCTGACGCGCCACAGACACTCCCTGGACGCGGCTGCTGGTG 351
Db 62 GluAspGlyGlyPheLeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuVal 81
QY 352 TACAAGGTGAAGATCCCTGGGCAACAACCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCG 411
Db 82 TyrLysValLysIleLeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLys 101
QY 412 GCGCGCGCTGGGAGCGCTCCACCGAGATCGTGTACGAGGTGGACGCGGCTGCTGGCGGCG 471
Db 102 ValGlyArgTrpGluProGlyThrGluIleValTyrGluValAspGlyValLeuArgGly 121
QY 472 CAGTCCCTGATGCCCTGGAGTCCCGGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 531
Db 122 GlnSerLeuMetAlaLeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThr 141
QY 532 ACCTACCGCTCCCAAGAGCGCGCTCCCGCTCAAGATGCCCGCTTCCACTTCGAGGAC 591
Db 142 ThrTyrArgSerLysLysProAlaSerAlaLeuLysMetProGlyPheHisPheGluAsp 161
QY 592 CACCGCATCGAGATCTCGGAGGAGGTGGAGAGGGCAAGTGTCTACAGCAGTACGAGGCG 651
Db 162 HisArgIleGluIleMetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAla 181
QY 652 GCGGTGGCGCGCTACTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699
Db 182 AlaValGlyArgTyrCysAspAlaAlaProSerLysLeuGlyHisAsn 197

RESULT 13
US-10-724-178-1065
; Sequence 1065, Application US/10724178
; Publication No. US20040137528A1

```

GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1065
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.: KFP F1F, with position 1 Met removed
US-10-724-178-1065

Alignment Scores:
Pred. No.: 128-56 Length: 185
Score: 961.00 Matches: 176
Percent Similarity: 96.22% Conservative: 2
Best Local Similarity: 95.14% Mismatches: 7
Query Match: 73.92% Indels: 0
DB: 16 Gaps: 0

US-10-081-864-21 (1-707) x US-10-724-178-1065 (1-185)
QY 7 GCCTCCCTGTCGACGAGACCATGCTTCAGACACCATCGAGGACCGCTGAGACGCG 66
Db 1 AlaSerLeuLeuThrGluThrMetProPheLysThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTCACCGGCAAGCGGAGGCAACCCCTCGAGGSCACCCAGGAGATG 126
Db 21 HisCysPheLysCysIleGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGGTGATCGAGGCGCGCCCTGCTCCCTTCCTCCATCTCCACTTCAAG 186
Db 41 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGGCTCCAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db 51 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGCTTCACCTCGAGCGGACCTGCTGTGTACAGAGTGAAGATC 306
Db 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTrpGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACACGAGCACCTCCCTCGAGCGGACCTGCTGTGTACAGAGTGAAGATC 366
Db 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
QY 367 CTGGCAACAACCTCCCGCGGACCGCCCTGTATGACGAAACAGCGCGCGCTCGGAG 426
Db 121 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysValGlyArgTrpGlu 140
QY 427 CCCTCCACCGAGATCGTGTACGAGTGGACGGCTGTGCGCGGACCTGCTGTATGAGC 486
Db 141 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
QY 487 CTGGAGTCCCGCGGCTGCGCACTGACCTGACCTGACCTGACCACTGACCACTGAG 546
Db 161 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisThrThrThrArgSerLys 180
QY 547 AAGCCCGCTCCGCC 561
Db 181 LysProAlaSerAla 185

RESULT 14

US-10-724-178-1061
; Sequence 1061, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1061
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.: KFP F1E, with position 1 Met removed
US-10-724-178-1061

Alignment Scores:
Pred. No.: 228-50 Length: 168
Score: 868.00 Matches: 159
Percent Similarity: 95.83% Conservative: 2
Best Local Similarity: 94.64% Mismatches: 7
Query Match: 66.77% Indels: 0
DB: 16 Gaps: 0

US-10-081-864-21 (1-707) x US-10-724-178-1061 (1-168)
QY 7 GCCTCCCTGTCGACGAGACCATGCTTCAGACACCATCGAGGACCGCTGAGACGCG 66
Db 1 AlaSerLeuLeuThrGluThrMetProPheLysThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTCACCGGCAAGCGGAGGCAACCCCTCGAGGSCACCCAGGAGATG 126
Db 21 HisCysPheLysCysIleGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGGTGATCGAGGCGCGCCCTGCTCCCTTCCTCCATCTCCACTTCAAG 186
Db 41 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGGCTCCAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db 61 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGCTTCACCTCGAGCGGACCGACCATCGAGGACCGCGCTTC 306
Db 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTrpGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACACGAGCACCTCCCTCGAGCGGACCTGCTGTGTACAGAGTGAAGATC 366
Db 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 120
QY 367 CTGGCAACAACCTCCCGCGGACCGCCCTGTATGACGAAACAGCGCGCGCTCGGAG 426
Db 121 LeuGlyAsnAsnProAlaAspGlyProValMetGlnAsnLysValGlyArgTrpGlu 140
QY 427 CCCTCCACCGAGATCGTGTACGAGTGGACGGCTGTGCGCGGACCTGCTGTATGAGC 486
Db 141 ProGlyThrGluIleValTyrGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 160
QY 487 CTGGAGTCCCGCGGCTGCGCACTGACCTGACCTGACCACTGACCACTGAG 510
Db 161 LeuLysCysProGlyGlyArgHis 168

RESULT 15
US-10-724-178-1053

```
; Sequence 1053, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: MacDonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10724,178
; PRIOR FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: Patent version 3.0
; SEQ ID NO 1053
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KFP F1C, with position 1 Met removed
US-10-724-178-1053
```

```
Alignment Scores:
Pred. No.:      8.52e-45      Length:      152
Score:          785.00      Matches:     144
Percent Similarity: 95.39%      Conservative: 1
Best Local Similarity: 94.74%      Mismatches: 7
Query Match:      60.38%      Indels: 0
DB:              16      Gaps: 0
```

US-10-081-864-21 (1-707) x US-10-724-178-1053 (1-152)

```
QY 7 GCCTCCCTGTCGACGAGACATGCCCTTCAGGACACCATCGAGGCGACCGTGAAACGGC 66
Db 1 AlaSerLeuLeuThrGluThrMetProPheLysThrThrIleGluGlyThrValAsnGly 20
QY 67 CACTACTTCAAGTCCACCGGAGGCGAGGCAACCCCTCGAGGGCACCCAGGAGATG 126
Db 21 HisCysPheLysCysIleGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 40
QY 127 AAGATCGAGTGATCGAGGCGGCGCCCTTCCTTCCTTCACATCCTGTCACCTCC 186
Db 41 LysIleGluValIleGluGlyGlyProLeuProPheAlaPheHisIleLeuSerThrSer 60
QY 187 TGCATGTACGCGTCCAGGCGCTTCATCAGTAGTGTCTCGGCATCCCGACTACTTCAAG 246
Db 61 CysMetTyrGlySerLysThrPheIleLysTyrValSerGlyIleProAspTyrPheLys 80
QY 247 CAGTCCCTCCCGAGGCGCTTCACTGGGAGCGCACCCACCATCTACGAGGACGCGGCTTC 306
Db 81 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyPhe 100
QY 307 CTGACCGCCACACGACACCTCCCTGGAGCGGCACTGCTGGTGTAAGGTGAAGATC 366
Db 101 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValIle 120
QY 367 CTGGGCAACAACCTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
Db 121 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysValGlyArgTrpGlu 140
QY 427 CCTCCACCGAGATCGTGACAGGTGGACGCGGTG 462
Db 141 ProGlyThrGluIleValTyrGluValAspGlyVal 152
```

Search completed: July 29, 2004, 14:54:52
Job time : 47.9213 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:28:05 ; Search time 15.3141 Seconds

(without alignments)
8881.684 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 1300

Sequence: 1 ggatccgctcctcctgctgac.....ctgggccacaactgaagctt 707

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2 1/USPTO.spool.p/US10081864/runat 29072004 150506 25492/app_query.fasta_1.1742
-DB=pir_78 -QMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864 @CNG 1 1 53 -runat 29072004 150506 25492 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	257	19.8	3020	2 A43932	mucin 2 precursor,
2	249.5	19.2	464	2 S22697	extensin - Volvox
3	231.5	17.6	539	2 T28770	hypothetical prote
4	229	17.6	620	2 S06733	hydroxyproline-ric
5	222	17.1	580	2 T43481	probable mucin DKF
6	221.5	17.0	238	1 J01514	green-fluorescent
7	221.5	17.0	1188	2 S49915	extensin-like prot
8	218	16.8	760	2 F86387	probable Pto kinas
9	209	16.1	214	2 T09854	proline-rich cell
10	209	16.1	708	2 D96711	hypothetical prote
11	208	16.0	222	2 H96711	hypothetical prote
12	207	15.9	214	2 T10737	extensin-like cell
13	207	15.9	1201	2 G86441	unknown protein [i
14	203.5	15.7	351	2 S50754	hypothetical prote

15	202	15.5	461	2 T10741	extensin-like prot
16	202	15.5	473	2 S50755	hypothetical prote
17	199	15.3	350	2 S22456	hydroxyproline-ric
18	198.5	15.3	507	2 T44768	antifreeze glycope
19	198.5	15.3	786	2 T01456	extensin homolog F
20	196	15.1	328	2 JQ0985	hydroxyproline-ric
21	194.5	15.0	267	2 S08314	cell wall glycopro
22	194.5	15.0	279	2 S53363	mucin 5AC (clone J
23	193.5	14.9	303	2 S28264	hydroxyproline-ric
24	193.5	14.9	731	2 T04455	hypothetical prote
25	193.5	14.9	760	2 T06291	extensin homolog T
26	193	14.8	776	2 T01361	hypothetical 119.5
27	192	14.8	1106	2 JQ0405	hypothetical 119.5
28	189.5	14.6	707	2 A46302	prB-associated spl
29	189.5	14.6	839	2 T04859	extensin homolog F
30	189	14.5	228	2 S53504	extensin-like prot
31	189	14.5	377	2 A48018	mucin 7 precursor,
32	189	14.5	476	2 T27051	hypothetical prote
33	187	14.4	317	2 A28996	proline-rich prote
34	186.5	14.3	275	2 T51437	hypothetical prote
35	186.5	14.3	376	2 S71558	probable cell wall
36	186.5	14.3	1161	2 S57180	probable membrane
37	186	14.3	1151	2 T18535	high molecular mas
38	186	14.3	3534	2 T42567	tegument protein 2
39	185.5	13.8	319	2 F75420	hypothetical prote
40	184.5	14.2	264	2 S29893	salivary glue prot
41	184.5	14.2	393	2 P00479	pistil extensin-li
42	184.5	14.2	574	2 T43556	Miskott-Aldrich sy
43	184.5	14.2	574	2 T38819	wiskott-aldrich sy
44	184	14.2	191	2 F84522	probable prolina-r
45	184	14.2	216	2 I51920	mucin - rhesus mac

ALIGNMENTS

RESULT 1

A43932

mucin 2 precursor, intestinal - human (fragments)

N:Alternate names: mucin SMUC-41

C:Species: Homo sapiens (man)

C:Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 05-Nov-1999

C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, S.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the

A:Reference number: A49963; MUID:94132002; PMID:8300571

A:Accession: A49963

A:Molecule type: mRNA

A:Residues: 1-639 <GUI>

A:Cross-references: GB:L21998

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr

A:Reference number: A45106; MUID:93016075; PMID:1400449

A:Accession: A45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 626-1895 <GU2>

A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396

A:Note: sequence extracted from NCBI backbone (NCBIP:116706)

A:Accession: B45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 2037-3020 <GU3>

A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398

A:Experimental source: colon

A:Note: sequence extracted from NCBI backbone (NCBIP:116698)

R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M

J. Clin. Invest. 88, 1005-1013, 1991

A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp

A:Reference number: A43932; MUID:91358717; PMID:1885763

A:Accession: A43932

A:Molecule type: DNA

A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
 A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
 A:Note: sequence inconsistent with the nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIIP:55750)
 R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
 J. Biol. Chem. 264, 6480-6487, 1989
 A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evidence
 A:Reference number: A33532; MUID:89197956; PMID:2703501
 A:Accession: B33532
 A:Molecule type: mRNA
 A:Residues: 1916-2193 <GU4>
 A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
 A:Experimental source: intestine
 R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
 J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481; PMID:1985113
 A:Accession: A61257
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner, B.
 Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-
 A:Reference number: P00328; MUID:92198477; PMID:1550588
 A:Accession: P00328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M86523
 A:Experimental source: small intestine
 A:Accession: P00329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5-11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
 C:Keywords: Glycoprotein; intestine; tandem repeat
 F:2766-2834/Domain: von Willebrand factor type C repeat homology <WVC>
 Alignment Scores:
 Pred. No.: 1.56e-09 Length: 3020
 Score: 257.00 Matches: 83
 Percent Similarity: 39.93% Conservative: 24
 Best Local Similarity: 30.97% Mismatches: 93
 Query Match: 19.77% Indels: 68
 DB: 2 Gaps: 13
 US-10-081-864-21 (1-707) x A43932 (1-3020)
 QY 5 CCGCCT-----CCCTGCTGACGAGACCATGCGCTTCAGGACCAACCATCGAGG---- 52
 Db 1495 ProProThrThrThrProSerProProMetThrThrProleThrProProAlaSerThr 1514
 QY 53 -----GCACCGTGAACGGCCACTACT----- 73
 Db 1515 ThrThrLeuProProThrThrThrProSerProProThrThrThrThrThrProPro 1534
 QY 74 ---TCAAGTACCGGACGGGAGGAGGCAACCCCTCGAGGACCCAGGAGATGAAGA 130
 Db 1535 ProThrThrThrProSerProProThrThrThrProleThrProleThrProSerThr 1554
 QY 131 TCAGAGTGATCGAGGGCGCCCTGCGCTTCCTTCCACATCCCTGCTCCACTCTGCA 190
 Db 1555 ThrLeuProProThrThrThrProSerProProThrThrThrThrThrProProPro 1574
 QY 191 TGACGGCTCAAGGCTTCATCAAGTACGTGTCGGCATGCCGACT---ACTCAAGC 247
 Db 1575 ThrThrProSerProProThrThrThrProSerProProThrThrThrThrThr 1594
 QY 248 AGTCCCTCCCGAGGGGCTTCACCTGG-----AGCGCACCAACCATCT 289

Db 1595 ThrProProProThrThrThrProSerProProThrThrThrThrThrProProPro 1614
 QY 290 ACAGAGACGGCGCTTCCTGACGGCCACCGACACCT-----CCCTGGACGGCGACT 343
 Db 1615 ThrThrProSer-----ProProThrThrThrProleThrProProThrSerThr 1632
 QY 344 GCGTGTGTACAAAGGTGAAGATCCTGGGCAACAACTCCCGCCGACGCGCCCGGTGATGC 403
 Db 1633 -----ThrThrLeuProProThr----- 1638
 QY 404 AGAAGACGGCGCGCTGCGGAGCCCTCCACGAGATCGTGTACAGGTGACGCGGTGC 463
 Db 1639 ---ThrThrPro-----SerProProProThrThrThrThrThrProProThr 1654
 QY 464 TCGCGCGCGAGTCCCTGATGGCCCTGGAGTCCCGCGGGTGGCCACCTGACCTGCCACC 523
 Db 1655 ThrThrProSerPro-----ProThrThr 1662
 QY 524 TGCACA-----CCACTACCGCTCCAAAGACCGCCCTCCGCTCGAGATGCCCGCT 577
 Db 1663 ThrThrProSerProProleThrThrThrThrThrThrThrThrThrThrProSer 1682
 QY 578 TCACCTTCGAGGACCAACCGCATCGAGATCCTGGAGGAGTGGAGAGGCAAGTCTCTACA 637
 Db 1683 SerProleThrThrThrProSerProProThrThrThrThrThrThrThrThrProSerProThr 1702
 QY 638 AGCAGTACGAGCGCGCTGGCGCTACTCGAGCGCG----- 676
 Db 1703 ThrThrProSerProProleThrThrThrThrThrThrThrThrThrThrProSerSerThrThrThrProSer 1722
 QY 677 CCGCTTCAAGCTGGCGCAACT 700
 Db 1723 ProProProThrThrThrMetThrThr 1730
 RESULT 2
 S22697
 A:extensin - Volvox carteri (fragment)
 C:Species: Volvox carteri
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-Jul-2000
 C:Accession: S22697; S21006
 R:Erli, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
 EMBO J. 11, 2055-2062, 1992
 A:Title: A novel extensin that may organize extracellular matrix biogenesis in Volvox ca
 A:Reference number: S22697; MUID:92289669; PMID:1600938
 A:Accession: S22697
 A:Molecule type: mRNA
 A:Residues: 1-464 <HAL>
 A:Cross-references: EMBL:X65165; NID:g21991; PIDN:CAA46283.1; PID:g21992
 C:Keywords: glycoprotein
 Alignment Scores:
 Pred. No.: 4.87e-09 Length: 464
 Score: 249.50 Matches: 86
 Percent Similarity: 37.55% Conservative: 12
 Best Local Similarity: 32.95% Mismatches: 79
 Query Match: 19.19% Indels: 85
 DB: 2 Gaps: 13
 US-10-081-864-21 (1-707) x S22697 (1-464)
 QY 5 CCGCTCCCTGTCAGCGAGACCATGCCCTTCAGGACCACTCAAGG 52
 Db 223 ProProPro-----ProArgValSerThrSerProProProAlaArgValSerSer 240
 QY 53 GCACGGTGAACGGCGCATCTACTTCAGTGCACCGCGCAGG----- 91
 Db 241 SerProProAlaThrArgSerProProProArgIleThrSerProSerProVal 260
 QY 92 -----GCAGGGCAACCCCTTCGAGGCGCCAGGATGAAGA 130
 Db 261 LeuThrAlaSerProProLeuProLeuThrSerProProProPro----- 276

A:Accession: S06733
A:Molecule type: DNA
A:Residues: 1-620 <KE>
A:Cross-references: EMBL:X13885; NID:gl9866; PID:CAA32090.1; PID:gl9867
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: glycoprotein

Alignment Scores:
Pred. No.: 1,1e-07 Length: 620
Score: 229.00 Matches: 81
Percent Similarity: 40.23% Conservative: 22
Best Local Similarity: 31.64% Mismatches: 89
Query Match: 17.62% Indels: 64
DB: 2 Gaps: 11

US-10-081-864-21 (1-707) x S06733 (1-620)

```
QY 5 CCGCTCCCTGCTGACCGAGACCATGCCCTTCAGACCAACCATCGAGG----- 52
Db 207 ProProProThrHisValGlnProThrProSer---ProProSerArgGlyHisGlnPro 225
QY 53 -----GCACCGTGAACGGCCCACTTCAAGTGCACCGGCAAGG 91
Db 226 GlnProProThrHisArgHisAlaPro-----ProThrHisArgHisAlaProProThr 243
QY 92 GCGAGGGCAACCCCTCGAGG-----GCACCGAGGAGATGAAGATCGAGGTGA 139
Db 244 HisGlnProSerProLeuArgHisLeuProProSerProArgGlnProGlnProPro 263
QY 140 TCGAGGGGG-----GCCCGCTGCCCTTCGCTTCACATCC 175
Db 264 ThrTySerProProProProAlaTyAlaGlnSerProGlnProSerProProProPro 283
QY 176 TGTCCACTCTGCTGATGACGGCTTCAAGGCTTCAAGTACGTGTCGGCATCCCG 235
Db 284 ProProProProThrTySerProProProProSerProProProProProProPro 303
QY 236 ACTACTCAACGAGTCCCTCCCGGAGGCTTCACTGGGAGCGCACCACCACTACGAGG 295
Db 304 AlaTySerProSerProProProProProProProProProProProProProProPro 321
QY 296 ACGGCGGCTTCTGACCGCCACCGAGGACACTCCCTCGGACGGGACTGCTGTTGTA 355
Db 322 ProAlaTySerProProProProProProProProProProProProProProPro 333
QY 356 AGGTGAAGATCTGGGCAACACTTCCCGCGGAGCGCCCGTGTATCAGAACAGCGCG 415
Db 334 -----ProProProProProProProProProProProProProProProPro 344
QY 416 GCCGCTGGGAGCCCTCCACCGAGATCGTGTACGAGGTGGAGCGGCTGCTGCGGCGCAGT 475
Db 345 IleTySerProProProPro-----ValTySer 354
QY 476 CCTGTAGTGGCTGGAGTCCCGCGGTCGCCACCTGACCTGCGCAGTGCACCACT 535
Db 355 ProProProPro-----ProSerTySerProProProProProProProProProPro 371
QY 536 ACCGCTCAAGAACCGCTCGCCCTGAAGATGCGCGGCTCCACTTCGAGGACCA 595
Db 372 ProProProSerProProProProProProProProProProProProProProPro 391
QY 596 GCA-----TCGAGATCTGGAGGAGTGGAGAGGCAAGTGTCTACAGCACTACG 646
Db 392 SerProProProProProAlaTySerProProLeuProAlaProProProProProPro 411
QY 647 AGGCGCGCTGGCGGCTGCGACCGCCCGCCCTCCAGCTGGGCC 694
Db 412 ProProPro-----ThrTySerProProProProProProProProProProPro 424
```

RESULT 5

T43481
probable mucin DKFZp434C196.1 - human (fragment)
N:Alternate names: protein DKFZp434B0635.1

C:Species: Homo sapiens (man)
C>Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text_change 20-Apr-2000
C:Accession: T43481; T34549; T17264
R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z22514
A:Accession: T43481
A:Molecule type: mRNA
A:Residues: 1-580 <AAA>
A:Cross-references: EMBL:AL133561; NID:g6599133; PID:CA63715.1; PID:g6599134
A:Experimental source: adult testis; clone DKFZp434C196
R:Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A:Reference number: Z21540
A:Accession: T34549
A:Molecule type: mRNA
A:Residues: 262-580 <POU1>
A:Cross-references: EMBL:AL122069; NID:g6102864; PID:CA659245.2; PID:g7018420
A:Experimental source: adult testis; clone DKFZp434B0635
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: Z18723
A:Accession: T17264
A:Molecule type: mRNA
A:Residues: 262-580 <POU2>
A:Cross-references: EMBL:AL117481; NID:gs911958; PID:CA655954.1; PID:gs911959
A:Experimental source: adult testis; clone DKFZp434B061
C:Genetics:
A:Note: DKFZp434C196.1; DKFZp434B0635.1; DKFZp434B061.1

Alignment Scores:
Pred. No.: 3,2e-07 Length: 580
Score: 222.00 Matches: 94
Percent Similarity: 39.55% Conservative: 12
Best Local Similarity: 35.07% Mismatches: 95
Query Match: 17.08% Indels: 68
DB: 2 Gaps: 14

US-10-081-864-21 (1-707) x T43481 (1-580)

```
QY 20 CCGAGACCATGCCCTTCAGGACCAACCATCGAGGCA-----CCGTGA 61
Db 65 ProAlaSerLeuMetArgThrProThrArgAlaSerLeuMetArgThrProProArg 84
QY 62 ACGGCCACTACTTCAAGTGCACCGGACGGGAGGCAACCCCTCGAGGCA---CCC 118
Db 85 AlaSerProThrArgLysProProArgAlaSerProArgThrProSerArgAlaSerPro 104
QY 119 AGGAGATGAAGATCGAGGTGATCGAGGCGGCGCCCTCGCCCTTCACATCCTGT 178
Db 105 ThrArg-----ArgLeuProArgAlaSerProMetGlySerProHisArgAla--- 120
QY 179 CCACCTCTGTGATGATCGGCTCCAGGCTTCATCAAGTGTGTCGGGATCCCGACT 238
Db 121 ---SerProMetArgThrProProArg-----AlaSerProThr 132
QY 239 ---ACTTCAAGCAGTCCCTCCCGAGGCTTCACTGGGAGCGCACCACTACGAGG 295
Db 133 GlyThrProSerThrAlaSerProThrGlyThrProSerSerAla---SerProThrGly 151
QY 296 ACGGCGGCTTCTGACCGCCACCGAGGACCTCCCTGGAGCGGCGATGCTGTGTACA 355
Db 152 ThrProProArgAlaSerProThrGlyThrPro----- 163
QY 356 AGGTGAAGATCTGGGCAACACTTCCCGCGGAGCGCCCGTGTATGACAGAACAAAG-- 412
Db 164 -----ArgAlaTrpAlaThrArgSerProSerThrAlaSerLeuThrArgThrProSer 181
QY 413 -----CCGCGCGCTGGAGCCCT-----CCACCGAGATCG 442
Db 182 ArgAlaSerLeuThrArgTrpProProArgAlaSerProThrArgThrProProArgGlu 201
QY 443 TGTACGAGGTGAGCGGCGTGTGTCGGCGGCGCAGTCCCTGATGGCCCT---GGAGTGCCTCCG 499
```

Db 202 SerProArgMetSer-HisArgAlaSerProThrArgThrProProArgAlaSerProTh 221
QY 500 GCGGTGCCACCTGACCTGCCACCTGCACACACCTACCGCTCCAAAGAGCCCGCTCG 559
Db 221 xArgArgProProArgAlaSerProThrArgThrProProArgGluSer-----LeuAr 239
QY 560 CCTGAGATGCCCGCTTCCACTTCGAGGACACCGCATCGAGATCCTGG-AGGAGTGG 618
Db 239 gThrSerHisArgAlaSerProThrArgMetProProArgAlaSerProThrArgArgPr 259
QY 619 GAGAGGCGCAAGTGCCTACAGCAGT-----ACGAGCCCGCTGGGCG 660
Db 259 oProArgAlaSerProThrGlySerProProArgAlaSerProMetThrProProArgAl 279
QY 661 CGCTACTGCGACCGCCGCC----- 679
Db 279 aSerProArgThrProProArgAlaSerProThrThrThrProSerArgAlaSerLeuTh 299
QY 680 ----CCTCCAAAGCTGGGCCACA 697
Db 299 rArgThrProSerTrpAlaSer 306

RESULT 6
JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001
C:Accession: JS0892; JQ1514; FQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Friedberg, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JQ1514; MUID:92175527; PMID:1347277
A:Accession: JS0692
A:Molecule type: DNA
A:Residues: 1-107, 'S', 109-238 <PRA1>
A:Cross-references: GB:M62654; NID:G155662; PIDN:AAA27722.1; PID:G155663
A:Accession: JQ1514
A:Molecule type: mRNA
A:Residues: 1-99, 'F', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>
A:Cross-references: GB:M62653; NID:G155660; PIDN:AAA27721.1; PID:G155661
A:Accession: FQ0335
A:Molecule type: protein
A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>
R:Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>
A:Cross-references: GB:L29345; NID:G606383; PIDN:AAA58246.1; PID:G606384
R:Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13, 'V', 15-24, 'Q', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'R',
A:Cross-references: EMBL:X83959; NID:G634008; PIDN:CAA58789.1; PID:G634009
A:Experimental source: clone gfp1
A:Accession: S51331
A:Molecule type: mRNA
A:Residues: 1-24, 'Q', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'Q',
A:Cross-references: EMBL:X83960; NID:G634010; PIDN:CAA58790.1; PID:G634011
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65692; PDB:1GLT
A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 81-9
A:Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996

A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A58953; MUID:98294543; PMID:9631087
A:Contents: annotation: X-ray crystallography, 1.9 angstroms
C:Comment: This protein is excited by the photoprotein aequorin (see PIR:AGJFNV) emitting
C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
C:Genetics:
A:Gene: GFP
A:Introns: 69/3; 167/3
C:Superfamily: green-fluorescent protein
C:Keywords: chromoprotein; luminescence
F:65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Alignment Scores: 3.45e-07 Length: 238
Pred. No.: 221.50 Matches: 60
Score: 47.00% Conservative: 42
Percent Similarity: 27.65% Mismatches: 78
Best Local Similarity: 17.04% Indels: 38
Query Match: 1 Gaps: 7
DB: 1

US-10-081-864-21 (1-707) x JQ1514 (1-238)

QY 13 CTGCTGACGAGACCATGCTTCAGGACACCATCGAGGACCGTGAACGGCCACTAC 72
Db 7 LeuPheThrGlyValValProIleLeuValGluLeuAspGlyAspValAsnGlyHisLys 26
QY 73 TTCAAGTCACCGGAGGCGAGGCGAACCCCTCGAGGCGACCCAGGAGATGAAGATC 132
Db 27 PheSerValSerGlyGluGlyAspAlaThrTyrGlyLysLeuThrLeuLysPhe 46
QY 133 GAGGTGATCGAGGCGGCGCCCTCGCTCCACATCCCTGCTCCACCTCCCTGTCATG 192
Db 47 ---IleCysThrThrGlyLysLeuProValProThrProThrLeuValThrPheSer 65
QY 193 TACGCTCCAAAGCGCTTCATCAAGTACGTGTCGGCATCCCTCC---GACTACTTCAAG 246
Db 66 TyrGlyValGlnCysPheSerArgTyrProAspHisMetLysGlnHisAspPhePheLys 85
QY 247 CAGTCCCTCCCGAGGCTTACCTGGGAGGCGACCCACCTACGAGGCGGCGCTTC 306
Db 86 SerAlaMetProGluGlyTyrValGlnGluArgThrIlePheTyrLysAspGlyAsn 105
QY 307 CTGACCGCCACCGAGACACCTCCCTCGAGCGGCGACTGCTGGTGTACAAAGTGAAGATC 366
Db 106 TyrLysThrArgAlaGluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeu 125
QY 367 CTGGGCAACAACCTTCCCGCGAGCGCGCCCGTGTGACAGAACAG- 411
Db 126 LysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysMetGluTyrAsnTyr 145
QY 412 -----CCCGCGCTGGAGCCCTCCACCGAGATCGTGATC 447
Db 146 AsnSerHisAsnValTyrIleMetAlaAspLysGlnLysAsnGlyIleLysValAsnPhe 165
QY 448 GAGGTG-----GACGCGTGTCTGCCTCCGCG- 470
Db 166 LysIleArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsn 185
QY 471 ----CAGTCCCTGATGGCGCTGGAGTCCCGCGCGTCCACCTGACCTGCACCTGCA 527
Db 186 ThrProIleGlyAspGlyProValLeu-----LeuProAspAsn 198
QY 528 CACCACTACCGCTCCAAAGACCGCGCTCCCG----CCTGAGATGCCCGG 575
Db 199 HisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArg 215

RESULT 7
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence revision 12-May-1995 #text change 29-Oct-1999
C:Accession: S49915

R.Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
 submitted to the EMBL Data Library, June 1994
 A:Description: Pex genes: pollen-specific genes with extensin-like domains.
 A:Reference number: S49915
 A:Accession: S49915
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1188 <RUB>
 A:Cross-references: EMBL:234465; NID:5600117; PIDN:CAA84230.1; PID:g600118

Alignment Scores:
 Pred. No.: 3,46e-07 Length: 1188
 Score: 221.50 Matches: 68
 Percent Similarity: 38.59% Conservative: 25
 Best Local Similarity: 28.22% Mismatches: 73
 Query Match: 17.04% Indels: 75
 DB: 2 Gaps: 9

US-10-081-864-21 (1-707) x S49915 (1-1188)

QY 8 CCTCCCTGCTGACGAGACCATGGCT-----TCAGGACCAACCA----- 46
 Db 924 ProProMetThrProLysSerProProValValSerProProValValSerProProThr 943

QY 47 ---TCGAGGGCAGCGCTGAACCGCCACTACTTCAAGTCGACCGCAAGGGCGAGGCAAC 103
 Db 944 VallysSerProProAlaProValSerProProAlaThrProLysSerSer 963

QY 104 CCTCGAGGGACCCAGAGATGAAGATCGAGGTGTCGAGGGCGCCCTCCCTCCCTGCG 163
 Db 964 ProProProAlaProValAsnLeuProProProGluValLysSerProProProThr 983

QY 164 CTTCCACATCTGTCCACCTCTCGATGTACGGCTCAAGGCTTCAATCAAGTACGTGT 223
 Db 984 ProValSerProProAlaProLysSerProProProAlaProMetSerSer 1003

QY 224 CCGCATCCCGACTACTTCAAGC-----AGTCCCTCCCGAGGGTTCACCT----- 271
 Db 1004 ProProProGluValLysSerProProProAlaProValSerProProPro 1023

QY 272 -----GGGAGCGCACCAACCTACGAGGACGGCGCTTCTCAGCCGCCACGAGACA 325
 Db 1024 ProValLysSerProProProAlaProValSerProProProProValLysSer 1043

QY 326 CTTCCCTGGACGGCACTGCTGCTGCTGTAAGGTGAAGATCTCTGGGCAACACTTCCCG 385
 Db 1044 Pro-----Pro-----Pro 1045

QY 386 CCGAGCGCCCGTGATGCAGNACAGCGCGCGCTGGAGCCCTCCACGAGATCGTGT 445
 Db 1046 ProProAlaProValSerProProProValLysSerProProProAla--- 1064

QY 446 ACGAGGTGGACGGCTGCTGCGCGCCAGTCCCTGATGGCCCTGGAGTGGCCCGGGGTC 505
 Db 1065 -----ProLysSerProProPro 1071

QY 506 GCCACCTGACCTGCACCTGCACACCTACCGCTCCA---AGAGCCCGCTCCGCCC 562
 Db 1072 -----ProValLysSerProProProAlaProValSerProProProPro 1088

QY 563 TGAAGATGCCCGCTTCCACTTCGAGGACCAACCGCATCGAGATCTTGGAGGAGTGAGA 622
 Db 1089 VallysSerPro----- 1092

QY 623 AGGCAAGTGCTACAGCAGTACGAGCGCGCGTGGCCCTACTGCGAGCGCCCGCT 682
 Db 1093 -----ProProAlaProValSerProProProPro 1103

QY 683 CCA 685
 Db 1104 Pro 1104

RESULT 8

F86387
 probable Pto kinase interactor [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F86387
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chain, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719; PMID:11130712
 A:Accession: F86387
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-760 <STO>
 A:Cross-references: GB:AE005172; NID:gl1079512; PIDN:AAG29223.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Alignment Scores:
 Pred. No.: 5,89e-07 Length: 760
 Score: 218.00 Matches: 75
 Percent Similarity: 38.59% Conservative: 18
 Best Local Similarity: 31.12% Mismatches: 85
 Query Match: 16.77% Indels: 64
 DB: 2 Gaps: 11

US-10-081-864-21 (1-707) x F86387 (1-760)

QY 20 CCGAGACCATGCCCTTCAGGACCA-----CCATCAGGCGGACCGTGA 61
 Db 27 ProProGluProSerProGlyAspAsnAlaThrSerProThrArgGluPro--- 45

QY 62 ACGGCGCACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGCACCCAGG 121
 Db 46 -----ThrAsnGlyAsnProProGluThrThrAsnThrPro----- 57

QY 122 AGATGAAGATCAGGTGATCGAGGGCGCCCTCGCCCTTCGCCTTCACATCTGTCCA 181
 Db 58 -----AlaGlnSerProProGluThrProLeuSerSerPro 71

QY 182 CCTCTGTCATGTACGGCTCCAGGCGCTTCA-----TCAAGTACGTGTCCG 226
 Db 72 ProProGluProSerProProSerProSerLeuThrGlyProProProThrIlePro 91

QY 227 GCATCCCGCACTACTTCAAGCAGTCCCTCCCGAGGGCTTCACCTGGGAGCGCACCA 286
 Db 92 ValSerProProGluProSerProProPro---ProLeuProThrGluAlaProPro 110

QY 287 CTTACGAGAGCGGGCTTCTGACCGCCACCGACAGGACACCTCCCTGGACGCGACTGCC 345
 Db 111 ProAlaAsnProValSerProProGluSerProProProProProProProPro 128

QY 347 TGGTGTACAGGTGAAGATCTGGGCAACACTTCCCGCGCGAGCGCCCTGTATGCAGA 406
 Db 129 -----ProThrGluAlaProProThrProIleThrSer 140

QY 407 ACAAGCGCGCGGTGGAGGCGCTCCACCGAGATCGTGTACGAGTGCAGCGGTGTGC 466
 Db 141 ProSerProThrAsnProProProProProProProProProProProProPro 156

QY 467 GCGGCGACCTCCCTGATGCCCTGGAGTGCCTGGGCGCGGTGCGCCACCTGACCTGCCACCTGC 526
 Db 156 rLeuProAlaProAspProProSerAsnProLeuProProProLysLeuValProPro 176

QY 527 ACAC-----CACCTACCGCT-----CAGAAGAC-----CGGCTCGCGCTG 564
 Db 176 rHisSerProProArgHisLeuProSerProProAlaSerGluIleProProProAr 196

QY 565 AAGATCCCGGCTTCCACITTCGAGGACCAACCGCATCGAGATCTCGGAGGAGTGGAGAAG 624
 Db 196 gHisLeuProSerProPro----- 202
 QY 625 GGCAAGTGTACAAAGCAGTACGAGGCGCGTGGCGGCTACTTGCAGACGCGCCCTCC 684
 Db 203 -AlaSerGluArgProSerThrProProSerAspSerGluHisProSerProProPro 222
 QY 685 A 685
 Db 222 o 222

RESULT 9

T09854
 C:Species: Gossypium hirsutum (upland cotton)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C/Accession: T09854
 R/John, M.B.; Keller, G.
 A:Title: Characterization of mRNA for a proline-rich protein of cotton fiber.
 A:Reference number: Z16888; MUID:95334493; PMID:7610164
 A/Accession: T09854
 A/Status: preliminary; translated from GE/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-214 <JOH>
 A/Cross-references: EMBL:L47308; NID:9435038; PIDN:AAA79360.1; PID:9435039
 A/Experimental source: strain Coker 312; fiber
 C/Superfamily: proline-rich protein
 C/Keywords: cell wall

Alignment Scores:
 Pred. No.: 2,31e-06 Length: 214
 Score: 209.00 Matches: 75
 Percent Similarity: 39.57% Conservative: 16
 Best Local Similarity: 32.61% Mismatches: 69
 Query Match: 16.08% Indels: 70
 DB: 2 Gaps: 14

US-10-081-864-21 (1-707) x T09854 (1-214)

QY 32 CCTTCAGGACCAACATCGAGGACCGTGAACGGCCACTACTTCAAGTCACCGCAAGG 91
 Db 26 ProSerAsnProProThrSerThrPro-----AlaThrProThrProProAlaSer 42
 QY 92 GCGAGGCAACCCCTCGAGGACCCAGGAGATGAAGATCGAGTGCAGGGCGGCC 151
 Db 43 ThrProProProThrThrGlnAlaPro----- 51
 QY 152 CCTGCGCTTCGCTTCCACATCTCTGTCACCTCTGCATGTACGGCTCCAGGCTTCA 211
 Db 52 -----ProThrPro-----ThrAlaThrProProProValSerThrProPro-----ProThr 67
 QY 212 TCAAGTACGTGTCG-----GCATCCCGACTACTTCAAGCAGTCCCTCCCGAGGGCT 265
 Db 68 SerSerProProValThrAlaSerProProProValSerThrProProSerSer 87
 QY 266 TCACCTGGAGCGCACACACCTACGAGACGGCGGCTTCTGACCGCCACGAGACA 325
 Db 88 ProProAlaThrProProPro-----AlaSerProProProAlaThrPro 103
 QY 326 CCTCCCTGGAGCGGACTGCTGCTGTACAAGGTGAAGATCCTGGGCAACAACTCCCG 385
 Db 104 ProPro-----AlaSerPro 108
 QY 386 CCGACGGCCCGTGTATGAGAACAGGCGCGCTGGAGCCCTCCACCGAGATCGTGT 445
 Db 109 ProProAla-----ThrProProAlaSerProProPro----- 120
 QY 446 ACGAGTGGACGGCTGCTGGCGCGCAGTCCCTGATGGCCCTGGAGTCCCGCGGCGTC 505
 Db 121 -----AlaThrPro-----ProProAlaThrProProPro 130

QY 506 GCCACCTGACCTGCCACCTGCACACACCACTACCGTCCCAAGAGCCCGCTCCG----- 559
 Db 131 AlaThr-----ProProProAlaThrProProProAlaProLeuAlaSerProProAlaThr 149
 QY 560 ---CCCTGAAGATGCCCGCTTCCACTTCGAGGACCCCGCATCGAGATCTCGAGGAGG 616
 Db 150 ValProAlaIleSerProValGlnThrProLeuThrSerProPro----- 164
 QY 617 TGGAGAGGGCAAGTGTACAAAGCAGTACGAGGCGCGCTGGCGGCTACTTGCAGCGCG 676
 Db 165 -----AlaProProThrGluAlaProAlaProThrLeuGlyAla---AlaThrPro 180
 QY 677 CCCCCTCAAGCTGGGCCAACACTGAAGCT 706
 Db 181 GlyProAlaGlyThrAspThrSerGlyAla 190

RESULT 10

D96711
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C/Accession: D96711
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.,
 ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: AB6141; MUID:21016719; PMID:11130712
 A/Accession: D96711
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-708 <STO>
 A/Cross-references: GB:AE005173; NID:95734709; PIDN:AAD49974.1; GSPDB:GNO0141
 C/Genetics:
 A/Gene: F24J5.8
 A/Map position: 1

Alignment Scores:
 Pred. No.: 2,32e-06 Length: 708
 Score: 209.00 Matches: 92
 Percent Similarity: 36.75% Conservative: 19
 Best Local Similarity: 30.46% Mismatches: 106
 Query Match: 16.08% Indels: 85
 DB: 2 Gaps: 14

US-10-081-864-21 (1-707) x D96711 (1-708)

QY 5 CCGCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCA----- 55
 Db 33 ProProValThrSerProLeuProProSerAlaProProProAsnAlaProProPro 52
 QY 56 CCGTGAACGGCCACTACTTCAAGTGCACCGCAAGGCGGAGGCAACCCCTCGAGGCA 115
 Db 53 ProProProValThrThrSerProProProValAlaAsnGlyAlaProProProLeu 72
 QY 116 CCACGAGATGAAGATCGAGG-----TGATCGAGGCGGCCCTTCGCTTCGCT 166
 Db 73 ProLysProProGluSerSerSerProProProGlnProValIleProSerProPro 92
 QY 167 TCACATCTCTGCCACCT---CCTGCATGTACGGCTCCAGGCTTCATCAAGTACGTGT 223
 Db 93 SerThrSerProProProGlnProValIleProProProProProProProProPro 112
 QY 224 CCGGCA-----TCCCGACTTACTCAAGCAGTCCCTCCCG 259
 Db 113 ProAlaLeuValProProLeuProProSerProProProProAlaSerValProProPro 132

Qy	626	GCAAGTGTACAA	AGCAGTACGAGGCGCGCTACTGCGACGCGCCCTCCA	685
		:::	:::	
Db	221	-----AlaTyrThrGlnProPhe-----	-----LysThrProProPro	233

Search completed: July 29, 2004, 14:37:37
Job time : 22.3141 secs

Blank Sheet - US PTD

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:24:20 ; Search time 10.2094 Seconds
(without alignments)
7211.716 Million cell updates/sec

Title: US-10-081-864-21
Perfect score: 1300
Sequence: 1 ggatcgctccctgctgac.....ctgggcacacaaagagctt 707

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q/cgn2 1/USPTO spool p/US10081864/runat 29072004 150505 25463/app query.fasta_1.1742
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -ICOPCL=0 -ICOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864@cgn 1.1.22 @runat 29072004 150505 25463 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	257	19.8	5179	MUC2_HUMAN	Q02817 homo sapien
2	239.5	18.4	555	GP1_CHLRE	Q9fpq6 chlamydomon
3	229	17.6	620	EXTN_TOBAC	P13983 nicotiana t
4	217.5	16.7	238	GFP_AEQVI	P42212 aequorea vi
5	201.5	15.5	1033	IF2_STRCO	Q8cjg8 streptomyc
6	194.5	15.0	267	EXTN_MAIZE	P14918 zea mays (m
7	193	14.8	279	Y091_NPVP	O10341 orgyia pseu
8	189.5	14.6	707	SFPQ_HUMAN	P23246 homo sapien
9	186.5	14.3	1161	DAN4_YEAST	P47179 saccharomyc
10	183.5	14.1	283	EXTN_SORBI	P24152 sorghum bic
11	183.5	14.1	296	PMP3_MOUSE	P05143 mus musculu
12	183.5	14.1	1367	AMYH_YEAST	P08640 saccharomyc
13	182.5	14.0	1255	MUC1_HUMAN	P15941 h mucin 1 p
14	182	14.0	297	SGS4_DROME	Q00725 drosophila
15	182	14.0	1046	IF2_STRAW	Q02k53 streptomyc
16	179.5	13.8	802	ENAH_MOUSE	Q03173 mus musculu
17	178	13.3	1733	VNUA_PVKA	P33485 pseudorabie
18	176.5	13.6	1233	MUSA_HUMAN	P98088 homo sapien

19	175	13.5	964	1	IF2_COREF	Q8fpa7 corynebacte
20	174	13.0	1033	1	IF2_SURCO	Q8cjg8 streptomyc
21	172	13.2	860	1	YHLI_EBV	P03181 epstein-bar
22	170.5	13.1	263	1	ICP3_HSV1F	P08353 herpes simp
23	170	13.1	1125	1	IF2_PROMM	Q7vsm4 prochloroco
24	169.5	13.0	699	1	VGLG_HSV2H	P13290 herpes simp
25	169	13.0	261	1	PRP2_MOUSE	P05142 mus musculu
26	169	13.0	426	1	EXLP_TOBAC	Q03211 nicotiana t
27	169	13.0	670	1	SYN1_MOUSE	Q89935 mus musculu
28	169	13.0	2715	1	MLL4_HUMAN	Q9unn6 homo sapien
29	168.5	13.0	1009	1	IF2_CAUCR	Q9ac25 caulobacter
30	166.5	12.8	534	1	APG_ARATH	P40602 arabidopsis
31	166.5	12.8	817	1	VRP1_YEAST	P37370 saccharomyc
32	166.5	12.8	5703	1	MUSB_HUMAN	Q9hnc84 homo sapien
33	166	12.4	1046	1	IF2_STRAW	Q8k53 streptomyc
34	165.5	12.7	1185	1	DRF_HUMAN	P54259 homo sapien
35	165	12.7	704	1	SYN1_RAT	P09951 rattus norv
36	164	12.6	437	1	YGY3_HALSQ	P21561 haloferax s
37	164	12.6	3149	1	TEGU_EBV	P03186 epstein-bar
38	163.5	12.6	1004	1	IF2_CORGL	Q8mp40 corynebacte
39	163	12.5	331	1	PRE1_HUMAN	Q04280 homo sapien
40	162.5	12.5	503	1	WAP1_HUMAN	O43516 homo sapien
41	162.5	12.5	1003	1	MD6_HUMAN	Q96dn6 homo sapien
42	162	12.5	400	1	MUAI_XENLA	P10667 xenopus lae
43	161	12.4	505	1	WASL_BOVIN	Q95107 bos taurus
44	161	12.0	641	1	EBN1_EBV	P03211 epstein-bar
45	161	12.4	917	1	SMOO_HUMAN	P53814 homo sapien

ALIGNMENTS

RESULT 1
MUC2_HUMAN STANDARD; PRT; 5179 AA.
ID MUC2_HUMAN Q02817; Q14878;
AC 002817; Q14878; (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
OS MUC2 OR SMUC.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Roche E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and
CC other mucus membrane-containing organs. Thought to provide a
CC protective, lubricating barrier against particles and infectious

FT STRAND 92 100
 FT TURN 101 102
 FT STRAND 105 115
 FT TURN 116 117
 FT STRAND 118 128
 FT TURN 132 133
 FT STRAND 135 139
 FT TURN 141 141
 FT STRAND 148 155
 FT TURN 156 159
 FT STRAND 160 171
 FT TURN 172 173
 FT STRAND 176 187
 FT TURN 199 208
 FT STRAND 211 212
 FT TURN 217 227
 SQ SEQUENCE 238 AA; 26986 MW; EA5A6P21PFB6E05 CRC64;

Alignment Scores:
 Pred. No.: 3,43e-05 Length: 238
 Score: 217.50 Matches: 59
 Percent Similarity: 47.00% Conservative: 43
 Best Local Similarity: 27.19% Mismatches: 78
 Query Match: 16.73% Indels: 38
 DB: 1 Gaps: 7

US-10-081-864-21 (1-707) x GFP_AEQVI (1-238)

QY 13 CTGTGACGAGACCATGCTTCAGGACCAACCATCGAGGACCCGCGTGAACGCCACTAC 72
 Db 7 LeupheThrGlyValProileuValGluLeuAspGlyAspValAsnGlyHisLys 26
 QY 73 TTCAGTCCACCGGAGGCGGAGGCAACCCCTCGAGGCGCCAGGAGATGAATC 132
 Db 27 PheSerValSerGlyGluGlyGluGlyAspAlaThrTyrGlyLysLeuThrLeuLysPhe 46
 QY 133 GAGGTGATGAGGCGGCGGCTTCCTGCTCCATCCATCCCTGCTGATGCTGATG 192
 Db 47 ---lleCysThrThrGlyLysLeuProValProThrProThrLeuValThrThrPheSer 65
 QY 193 TAGGCTCCAGGCTTCATCAAGTACGTGTCGGCATGCC---GACTACTTCAAG 246
 Db 66 TyrGlyValGlnCysPheSerArgTyrProAspHisMetLysGlnHisAspPheLys 85
 QY 247 CAGTCCCTCCCGAGGCTTCACCTGGGAGGACCAACCATCGAGGCGGCGGCTTC 306
 Db 86 SerAlaMetProGluGlyTyrValGlnGluThrThrLeuPhePheLysAspGlyAsn 105
 QY 307 CTGACCGCCACCGACCATCCCTGAGACGCGGACTGCTGTGTACAGGTGAAGATC 366
 Db 106 TyrLysThrArgAlaGluValLysPheGluGlyAspThrLeuValAsnArgIleGluLeu 125
 QY 367 CTGGCACACACTTCCCGCGGAGGCGGCGGCTGTGATGACAGACAG--- 411
 Db 126 LysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHisLysLeuGluTyrAsnTyr 145
 QY 412 -----GCCGCGCGTGGGAGCGCTCCACCGAGATCGTGTAC 447
 Db 146 AsnSerHisAsnValTyrIleValAlaAspLysGlnLysAsnGlyIleLysValAsnPhe 165
 QY 448 GAGGTG-----GAGGCGGTGTCGCGCG--- 470
 Db 166 LysIleArgHisAsnIleGluAspGlySerValGlnLeuAlaAspHisTyrGlnGlnAsn 185
 QY 471 ---CCAGTCCCTGATGGCTGAGTGCCTCGGCGGCGGCTGACCTGACCTGCA 527
 Db 186 ThrProIleGlyAspGlyProValLeu-----LeuProAspAsn 198
 QY 528 CACACCTACGCTCCAGAGAGCGGCTCCGCG---CCTGAGATGCCCGG 575
 Db 199 HisTyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArg 215

RESULT 5

IF2 STRCO
 ID IF2 STRCO STANDARD; PRT; 1033 AA.
 AC Q8CJQ8;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Translation initiation factor IF-2.
 GN INFB OR SC05706 OR SC9F2.10C OR SC5H4.30.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]_TaxID=1902;
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=2196410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)";
 RL Nature 417:141-147(2002).
 CC -!- FUNCTION: One of the essential components for the initiation of
 CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
 CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
 CC Also involved in the hydrolysis of GTP during the formation of the
 CC 70S ribosomal complex (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the IF-2 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL939124; CAD55362.1; -.
 DR PIR; T35989; T35989.
 DR HAMAP; MF_00100; -; 1.
 DR InterPro; IPR000637; AT_hook.
 DR InterPro; IPR001851; Bac_inmem_transp.
 DR InterPro; IPR004161; EFTU_D2.
 DR InterPro; IPR000795; EF_GTPbind.
 DR InterPro; IPR000178; IF2.
 DR InterPro; IPR006847; IF2_N.
 DR InterPro; IPR001806; Ras_transfmg.
 DR InterPro; IPR005225; Small_GTP.
 DR InterPro; IPR009000; Translat_factor.
 DR Pfam; PF03144; GTP_EFTU; 1.
 DR Pfam; PF04760; IF2_N; 2.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR PRINTS; PR00929; ATHOOK.
 DR PRINTS; PR00315; ELONGATNFACT.
 DR PRINTS; PR00449; RASTRNSFRMNG.
 DR ProDom; PD186100; IF2; 1.
 DR TIGRFAMs; TIGR00487; IF-2; 1.
 DR TIGRFAMs; TIGR00231; small_GTP; 1.
 DR PROSITE; PS01176; IF2; FALSE NEG.
 KW Initiation factor; Protein biosynthesis; GTP-binding;
 KW Complete proteome.
 FT DOMAIN 529 681 G-DOMAIN
 FT NP_BIND 535 542 GTP (BY SIMILARITY).
 FT NP_BIND 585 589 GTP (BY SIMILARITY).
 FT NP_BIND 639 642 GTP (BY SIMILARITY).
 SQ SEQUENCE 1033 AA; 105657 MW; 47D581FB0072A045 CRC64;

Db 222 oGlyLeuSerThrProGlyGlyHisProLysProHisArgGlyGlyGluProAr 242
 QY 525 -----GCACACACCTACCGCTCCAGAGAGCGCGCTCCGCTGAGATGCCG 574
 Db 242 gGlyArgGlnHisHisProProTyHisGlnGlnHis ----- 255
 QY 575 GCTTCCACTTCGAGGACACCGCATCGAGATCCTGGAGGAGGTGAGAGAGGCAAGTGCT 634
 Db 256 -----HisGlnGlyProProGlyGlyProGlyGlyArgSerGluGluLysIleSe 273
 QY 635 ACAACGACATAGAGCGCGCGCTGGCGCTACTGCGAGCGCGCC 680
 Db 273 rAspSerGluGlyPheLysAlaAsnLeuSerLeuLeuArgArgPro 288

RESULT 9
 ID DAN4_YEAST STANDARD; PRT; 1161 AA.
 AC P47179;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cell wall protein DAN4 precursor.
 GN DAN4 OR YJ151C OR J2223.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Scarcez T.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REGULATION.
 RX MEDLINE=21113168; PubMed=11160904;
 RA Cohen B.D., Sertil O., Abramova N.E., Davies K.J., Lowry C.V.;
 RT "Induction and repression of DAN1 and the family of anaerobic
 RT mannoprotein genes in saccharomyces cerevisiae occurs through a
 RT complex array of regulatory sites."
 RL Nucleic Acids Res. 29:799-808(2001).
 CC -!- FUNCTION: Component of the cell wall (By similarity).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (potential).
 CC -!- PTM: Extensively O-glycosylated (Potential).
 CC -!- SIMILARITY: Belongs to the SRP1 / TIP1 family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z49651; CAA89684.1; -;
 DR PIR; S57180; S57180.
 DR GenOnline; 141984; -;
 DR SGD; S0003912; DAN4.
 DR GO; GO:0009277; C:cell wall (sensu Fungi); IDA.
 DR InterPro; IPR000992; SRP1_TIP1.
 DR Pfam; PF00660; SRP1_TIP1; 1.
 DR PROSITE; PS00724; SRP1_TIP1; 1.
 KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal; Lipoprotein.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 1146 CELL WALL PROTEIN DAN4.
 FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
 FT LIPID 1146 1146 GPI-anchor amidated asparagine
 FT (potential).
 SQ SEQUENCE 1161 AA; 118358 MW; 7954C15D69F0CA58 CRC64;

Alignment Scores:
 Pred. No.: 0.0017 Length: 1161
 Score: 186.50 Matches: 61
 Percent Similarity: 42.27% Conservative: 32

Best Local Similarity: 27.73% Mismatches: 92
 Query Match: 14.35% Indels: 35
 DB: 1 Gaps: 5

US-10-081-864-21 (1-707) x DAN4_YEAST (1-1161)

QY 47 TCGAGGGCACCCTGAGCGCCACTTCAAGTCACCGGAGGCGGAGGCGCAACCC 106
 Db 150 SerThrProThrThrSerThrThrSerThrThrSerThrThrSerThrThrSer 168
 QY 107 TCGAGGGCACCAGGAGATGAGATCGAGGGCGCGCGCTGCGCTCGCT 166
 Db 169 -----ThrThrProThrThrSerThr 175
 QY 167 TCGACATCCTGCTCCACCTGCTGATGCTCCAGGCTTCAAGTACGTCGCT 226
 Db 176 ThrSerThrThrProThrThrSerThrThrProThrThrSerThrThrSer 195
 QY 227 GCATCCCGCCTACTTCA---AGCAGTCCCTCCCGAGGCTTACCTGGGAGCGCACCA 283
 Db 196 ThrThrProThrThrSerThrThrSerThrThrProThrThrSerThrThr 215
 QY 284 CCACCTAGGAGGAGCGCGCTTCCGACCGCCACGACACCTCCCTGGAGCGCGACT 343
 Db 216 ProThrThrSerThrThrSerThrThrProThrThrSerThrThrProThrThr 235
 QY 344 GCCTGGTGTACAGGTGAGATCTCTGGGCAACAACTCCCGCGGACGCGCGCTGATGC 403
 Db 236 ThrSerThrThrSerGlnThrSerThrLysSerThrThrProThrThrSerSer 254
 QY 404 AGAACAAGCGCGCGCTGGGAGCCCTCCACGAGATCGTGTACGAGTGGACGCGCTGC 463
 Db 255 SerThrThrProThrThrSerThrThrProThrThrSerThrThrSer 270
 QY 464 TCGCGGGCAGTCCCTGATGCGCTGGAGTGGCGGCTGCGGCTGCGCACCTGACCTGCCACC 523
 Db 271 -----ThrAlaProThrThrSerThrThrSerThrThr 281
 QY 524 TCGACACACCTACCGCTCCAGAGCGCGCTCGCGCTCGAGATGCGCGCTTCCACT 583
 Db 282 SerThrThrSerThrIleSerThrAlaPro-----ThrThrSerThr 295
 QY 584 TCGAGGACACCGCATCGAGATCCTGGAGGAGTGGAGAGGCGCAAGTGTACAGCAGT 643
 Db 296 ThrSerThrThrPheSerThrSerSerAlaSerAlaSerSerValIleSerThrAla 315
 QY 644 ACAGGCGCGCTGGCGCGCTACTGCGACGCGCGCGCTCCAGCTGGGCGCACACTGAA 703
 Db 316 ThrThrSerThrThrPheAlaSerLeuThrThrProAlaThrSerThrAlaSerThrAsp 335

RESULT 10
 EXTN_SORBI STANDARD; PRT; 283 AA.
 ID AC P24152;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Extensin precursor (Proline-rich glycoprotein).
 GN HRGP.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
 OX NCBI_TaxID=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RX MEDLINE=91370882; PubMed=1893107;
 RA Raz R., Cretin C., Puigdomenech P., Martinez-Izquierdo J.A.;
 RT "The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum
 RT vulgare."
 RL Plant Mol. Biol. 16:365-367(1991).
 CC -!- FUNCTION: Structural component in primary cell wall.

```

Db      196 -----ProValThrProSerProLysProLysProSerProLysProSer 208
Qy      665 -----ACTGCGAGCGCGCCCGCCCTCCCAAGC 688
Db      209 ProProThrThrProThrProLysProProAla 220

RESULT 11
PMP3 MOUSE
ID      PMP3_MOUSE STANDARD; PRT; 296 AA.
AC      P05143;
DT      13-AUG-1987 (Rel. 05, Created)
DT      13-AUG-1987 (Rel. 05, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Proline-rich protein MF-3 (Fragment).
GN      PRH1 OR PRP.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=86059475; PubMed=2999141;
RA      Ann D.K., Carlson D.M.;
RT      "The structure and organization of a proline-rich protein gene of a
RT      mouse multigene family.";
RL      J. Biol. Chem. 260:15863-15872(1985).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M12100; AAA40005.1; --
DR      MGD; MGI:97773; Pth1.
FT      NON TER.
SQ      SEQUENCE 296 AA; 29521 MW; 7F146824E8AF3269 CRC64;

Alignment Scores:
Pred. No.: 0.00255 Length: 296
Score: 183.50 Matches: 72
Percent Similarity: 35.86% Conservative: 13
Best Local Similarity: 30.38% Mismatches: 93
Query Match: 14.12% Indels: 60
DB: 1 Gaps: 12

US-10-081-864-21 (1-707) x PMP3_MOUSE (1-296)
Qy      23 AGACCATGCGCCTTCAGAGCACCATGAGCGGACCGGTGAACGGCCACTACTTCAAGTGCA 82
Db      12 ArgProProSerGlySerGlnProArgProProValAsnGlySerGlnGlnGlyPro 31
Qy      83 CCGGCAAGGCGCAGGG-----CAACCCCTCGAGGGCACCCA-----GGAGA 124
Db      32 ProProProGly-GlyProGlnProArgProProGlnGlyProProProGlyGlyPr 51
Qy      125 TGAAGATCGAGGTGATCGAGGGCGGCCCTGCG-----CCTTCGCTTCCAC 171
Db      51 oGlnProArgProProGlnGlyProProProGlyGlyProGlnProArgProProG 71
Qy      172 ATCCTGTCCACCTTCCTGCATGTAGCGCTCAAGGCTTCATCAAGTACGTGTCCGCGATC 231
Db      71 nGlyProProProProGlyGlyProGlnProArgPro-----ProGlnG 86
Qy      232 CCGCACTACTTCAAGCAGTCCCTCCCGGGGTTCACCTGGAGGCGCACCAACCTAC 291
Db      86 yProProProProGlyGlyProGlnProArg-----ProProGlnGlyProProPro 104
Qy      292 GAGGAGGGGGCTTCTGTACCGCCGCCACAGGACACCTCCCTGGACGGCACTGCCTGGT 351

```


RL Int. J. Cancer 66:55-59 (1996).
RN [15]
RN SEQUENCE OF 1-89 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA Lee L.N., Luh K.T., Wu C.W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126 (1996).
RN [16]
RN SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Breast carcinoma;
RA Buluwa L., Liu O., Luqmani Y.A., Gomm J.J., Coombes R.C.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
RN [17]
RN CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=97460054; PubMed=9312074;
RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
RA Hanisch F.-G.;
RT "Localization of O-glycosylation sites on glycopeptide fragments from
RT lactation-associated MUC1. All putative sites within the tandem
RT repeat are glycosylation targets in vivo.";
RL J. Biol. Chem. 272:24780-24793 (1997).
RN [18]
RN CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=99303572; PubMed=10373415;
RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
RA Hanisch F.-G.;
RT "High density O-glycosylation on tandem repeat peptide from secretory
RT MUC1 of T47D breast cancer cells.";
RL J. Biol. Chem. 274:18165-18172 (1999).
RN [19]
RN POLYMORPHISM WITHIN THE REPEAT.
RX MEDLINE=21359366; PubMed=11350974;
RA Engelmann K., Balduz S.B., Hanisch F.-G.;
RT "Identification and topology of variant sequences within individual
RT repeat domains of the human epithelial tumor mucin MUC1.";
RL J. Biol. Chem. 276:27764-27769 (2001).
RN [20]
RN CHARACTERIZATION OF ISOPORM Y, AND MUTAGENESIS OF ASP-1116.
RX MEDLINE=99211485; PubMed=10197628;
RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,
RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
RA Wreschner D.H.;
RT "The breast cancer-associated MUC1 gene generates both a receptor and
RT its cognate binding protein.";
RL Cancer Res. 59:1552-1561 (1999).
RN [21]
RN PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
RX MEDLINE=21240104; PubMed=11341784;
RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
RA Harris A.,
RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
RL Biochem. Biophys. Res. Commun. 283:715-720 (2001).
RN [22]
RN CHARACTERIZATION.
RX MEDLINE=21836452; PubMed=11847293;
RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,
RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
RA Stacey M., Lin H.-H., Gordon S.;
RT "Generation of ligand-receptor alliances by 'SEA' module-mediated
RT cleavage of membrane-associated mucin proteins.";
RL Protein Sci. 11:698-706 (2002).
RN [23]
RN PHOSPHORYLATION.
RX MEDLINE=95080414; PubMed=7988707;
RA Zrlhan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.;
RT "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.
RT Cytokine receptor-like molecules.";
RL FEBS Lett. 356:130-136 (1994).
CC -!- FUNCTION: May play a role in adhesive functions and in cell-cell
interactions, metastasis and signaling. May provide a protective

Alignment Scores: 0.00282 Length: 1255
Pred. No.: 182.50 Matches: 77
Score: 35.38% Conservative: 21
Percent Similarity: 27.80% Mismatches: 90
Best Local Similarity: 14.04% Indels: 89
Query Match: 1 Gaps: 14
DB: 1
US-10-081-864-21 (1-707) x MUC1_HUMAN (1-1255).
QY 23 AGACCATGCCCTTCAGGACCACCATCGAGGGGACCGTGAACGGCCACTTCAAGTGCA 82
Db 145 ArgProAlaProGly-----SerThrAlaProProAlaHisGlyValThrSerAla 161
QY 83 CCGGCAAGGGCGAGGGCAACCCCTCGAGGGGACCCAGGAGATGAAGATCGAGGTGTCG 142
Db 162 ProAspThrArgProAlaProGlySerThrAlaPro-----ProAlaHisGlyVal 178
QY 143 AGGGGGGGGGGG-----TGCCCTTCGCTTCCATCTCTGTCCACCT----- 184
Db 179 ThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHisGlyVal 198
QY 185 -----CCTGCATGTACGGCTCCA----- 202
Db 199 ThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHisGlyVal 218
QY 203 -----AGGCCTTCA-----TCAGTAGTGTCCGGCATCCCGGCTACT 241
Db 219 ThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHisGlyVal 238
QY 242 TCAAGCAGTCCCTCCCGAGGGCTTCACTGGGAGCGCACCCACACCTACGAGGACGGCG 301
Db 239 ThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHisGlyVal 258
QY 302 GTTCTGTACGCCCGCACGAGACACCTCCTGGAGCGGCGACTGCTGTGTACAGGTGA 361
Db 259 ThrSerAlaProAspThrArgProAlaPro----- 268
QY 362 AGNTCTGGGCAACAACCTCCCGCGG-----ACGGCCCGGTGATGTCAGA 406
Db 269 -----GlySerThrAlaProProAlaHisGlyValThrSerAlaPro-----Asp 283
QY 407 ACAAGCGCGCGCTGGGAGC-----CCTCACCGAGATCGTGTACGAGGTGGACGGCG 460
Db 284 ThrArgProAlaProGlySerThrAlaProProAlaHisGlyValThr----- 299
QY 461 TGTCTGGCGGCGAGTCCCTGTGAGTGGCGCGCGGCTCGCCACCTGACCTGCC 520
Db 300 -----SerAlaProAspThrArgProAlaProGly 309
QY 521 ACTGTGCACACACCT-----ACCGTCCCAAGAACCGCCCTCCGCGCTGA 565
Db 310 SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly 329
QY 566 AGATGCCCGGCTTCCACTTCGAGGACCGCATCGATCCTCGAGGAGGTGGAGAGG 625
Db 330 SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArg----- 345
QY 626 GCAAGTGTCTACAAGCAGTACGAGCGCGCGCTGGCGGCTACTCGACCGCGGCC----- 679
Db 346 ---ProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspThr 364
QY 680 -----CCTCAAGCTGGGCCACAACTGAAGCT 706
Db 365 ArgProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAla 381
RESULT 14
SGS4_DROME STANDARD; PRT; 297 AA.
ID SGS4_DROME
AC Q00725; Q9W4T2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Salivary glue protein Sgs-4 precursor.
 GN Sgs-4 OR EG96G10.6 OR CG12181.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R, Karsnas, and Samarkand;
 RX MEDLINE=92221133; PubMed=1562607;
 RA Furia M., Digilio F.A., Artiacono D., Favia G., Polito L.C.;
 RT "Molecular characterization of a Drosophila melanogaster variant
 strain defective in the Sgs-4 gene dosage compensation";
 RL Biochim. Biophys. Acta 1130:314-316(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
 RA Mirana B., Gloux S., Lelaou V., Mottier S., Galibert F., Borkova D.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo B.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourikoti F.,
 RA Beinert N., Dowe G., Schaefer U., Jaecle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlanitsou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valentini P., Saunders R.D.C.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of D.
 melanogaster";
 RL Science 287:2220-2222(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.W., Cawley S., Dahlke C., Davenport L.B., Davies F.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong P., Gorell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacieb J.M.,
 RA Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of Drosophila melanogaster";
 Science 287:2185-2195(2000).

-!- SUBCELLULAR LOCATION: Secreted (Potential).

-!- TISSUE SPECIFICITY: Salivary gland.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X61943; CAA43949.1; -

EMBL; X61942; CAA43948.1; -

EMBL; X61944; CAA43950.1; -

EMBL; AL024484; CAA19673.1; -

EMBL; AE003427; AAF45860.1; -

PIR; S21085; S21085.

DR FlyBase; FBgn0003374; Sgs4.

Repeat; Signal.

SIGNAL 1 21

CHAIN 22 297

DOMAIN 26 178

VARIANT 52 65

VARIANT 55 55

VARIANT 58 58

VARIANT 58 60

VARIANT 62 62

VARIANT 62 62

VARIANT 63 63

VARIANT 65 65

VARIANT 67 67

VARIANT 67 67

VARIANT 69 69

VARIANT 74 74

VARIANT 97 97

VARIANT 106 106

VARIANT 109 109

VARIANT 116 116

VARIANT 119 125

VARIANT 130 130

VARIANT 136 136

VARIANT 159 159

VARIANT 165 165

VARIANT 177 177

VARIANT 204 208

VARIANT 209 297

VARIANT 247 252

VARIANT 249 249

VARIANT 252 252

VARIANT 267 267

VARIANT 275 275

SEQUENCE 297 AA; 32309 MW; F121DDF2B177B58C CRC64;

Alignment Scores:

Pred. No.: 0.00308

Score: 182.00

Percent Similarity: 32.37%

Best Local Similarity: 26.26%

Query Match: 14.00%

DB: 1

US-10-081-864-21 (1-707) x Sgs4_DROME (1-297)

Length: 297

Matches: 73

Conservative: 17

Mismatches: 97

Indels: 91

Gaps: 15

```

QY 2 GATCGCCCTCCCTGTCACCGAGACCA-----TGCCCTTCAGGACCAACCACTCGAGGCA 55
Db 27 GluProProCysGluThrGluProProArgCysGluThrGluProPro----- 43
QY 56 CGGTGAACGGCCACTACTCAAGTCACGCGGAGGCGAGGCGACCCCTCGAGGCA 115
Db 44 -----ArgCysGluThrGluProProArgCysGluThrGluProProArg--- 58
QY 116 CCCAGGAGATGAAGATCGAGGTGTCGAGGCGCGCCCTCGCTCCCTCCATCATCC 175
Db 59 -----CysGluThrThrThrProLys 65
QY 176 TGT-----CACCTCCCTGCTAGTACGCTTCAGGCGCTTCATCAAGTACGTGCGGCA 229
Db 66 CysGluThrThrProProThrCysArgThrGluProPro-----ThrCysLysThr 82
QY 230 TCCCGGACTACTTCAAGCAGTCCCTCCCGAGGCGCTTCACTGGGAGCGCACCACT 289
Db 83 GluProProThrCysArgThrGluProProThrCysLys-----ThrLysProProThr 100
QY 290 ACGAGGCGCGGCTTCTGACCGCCACC---AGGACACTCCCTCGAGCGGCACTGCC 346
Db 101 CysArgThrGlu-----ProProThrCysArgThrGluProProThr----- 114
QY 347 TGGTGTACAGGTGAAGATCTGGGCAACACTTCCCGCGGAGCGCC----- 394
Db 115 -----CysLysThrLysProProThrCysLysThrGluPro 126
QY 395 CGGTGATGACAGAACGCGCGCGCTGGAGCCCTCCACCGAGATCGTGTACGAGTGG 454
Db 127 ProThrCysLysThrGluProProThrCysArgThrGluProProThrCysLysThrGlu 146
QY 455 ACGCGGTGCTCGGCGGCGGCTGCTGATGCGCCCTGGAGTGCCTCGCGGTGCGCACTGA 514
Db 147 ProProThrCysArgThrGluPro-----ProThrCysLysThrGlu 160
QY 515 CCTGCCACCTGC-----ACACCACTACCGCTCCAAAGAGCGCGCTCCGCGCTGAAGA 568
Db 161 ProProThrCysLysThrGluProProThrCysLysThrGluProProCysGluLysHis 180
QY 569 TGCCCGGCT-----TCCACTTCGAGGACCAACCGATCGAGA----- 604
Db 181 CysThrLysArgLysLysArgHisArgThrLysArgThrLysArgSerLysSerThrLys 200
QY 604 ----- 604
Db 201 LysileValHisHisAsnArgProGlyThrThrProGluSerGlyCysGlyCysGly 220
QY 605 -----TCCTCGAGAGGTGGAGAGGCGCAAGTGTCTACAAGCAGTACGAGCGCG 652
Db 221 SerLysAsnGluSer-GlyGlyGlySerGlyCysileLeuLysAspLeuLeuThrPr 240
QY 653 CCGTGGCGCGCTACTCGAGCGCGCGCGCTCCAGCTGGCGGCGGCACTGAAG 704
Db 240 oLysCysProAspSerLysProLysPro---GlnAlaSerProLysCysLys 256

RESULT 15
ID -IF2 STRAW STANDARD; PRT; 1046 AA.
AC Q82K53;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Translation initiation factor IF-2.
GN INF6 OR SAV2551.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;

```

```

RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the IF-2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF005031; BAC70262.1; -.
CC HAMAP; MF_00100; -.
CC DR InterPro; IPR001851; BAC_inmem_transp.
CC DR InterPro; IPR004161; EFTU_D2.
CC DR InterPro; IPR000795; EF_GTPbind.
CC DR InterPro; IPR003577; GTPase_Ras.
CC DR InterPro; IPR000178; IF2.
CC DR InterPro; IPR006847; IF2_N.
CC DR InterPro; IPR002365; P-rich_extensn.
CC DR InterPro; IPR001806; Ras_transfmg.
CC DR InterPro; IPR005225; Small_GTP.
CC DR InterPro; IPR009000; Translat_factor.
CC DR Pfam; PF00009; GTP_EFTU; 1.
CC DR Pfam; PF03144; GTP_EFTU_D2; 1.
CC DR Pfam; PF04760; IF2_N; 2.
CC DR PRINTS; PR00315; ELONGATNFCT.
CC DR PRINTS; PR01217; PRICHEXTNSN.
CC DR PRINTS; PR00449; RASTRNSFRNG.
CC DR PrDom; PD186100; IF2; 1.
CC DR SMART; SM00173; RAS; 1.
CC DR TIGRFAMs; TIGR00487; IF-2; 1.
CC DR TIGRFAMs; TIGR00231; small_GTP; 1.
CC DR PROSITE; PS01176; IF2; FALSE_NEG.
CC KW Initiation factor; Protein biosynthesis; GTP-binding;
CC Complete proteome.
CC FT DOMAIN 542 594 G-DOMAIN.
CC FT NP_BIND 548 555 GTP (BY SIMILARITY).
CC FT NP_BIND 598 602 GTP (BY SIMILARITY).
CC FT NP_BIND 652 655 GTP (BY SIMILARITY).
CC SQ SEQUENCE 1046 AA; 106882 MW; 62B1E2DFA3AA595 CRC64;
CC
CC Alignment Scores:
CC Pred. No.: 0.00301 Length: 1046
CC Score: 182.00 Matches: 86
CC Percent Similarity: 33.67% Conservative: 13
CC Best Local Similarity: 29.25% Mismatches: 94
CC Query Match: 14.00% Indels: 101
CC DB: 1 Gaps: 16
CC
CC US-10-081-864-21 (1-707) x IF2_STRAW (1-1046)

```

Search completed: July 29, 2004, 14:33:32
Job time : 18.2094 secs

Blank Sheet U.S. PTD

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 29, 2004, 14:27:25 ; Search time 42.1137 Seconds

(without alignments)
10593.763 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 1300

Sequence: 1 gggatccgctccctcgtgac.....ctggggccacactgaagctt 707

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2.1/USPTO.spool_p/US10081864/runat.29072004.150506.25475/app.query.fasta_1.1742
-DB=SPTREMBL 25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081864@cgn_1_122@runat.29072004.150506.25475 -NCPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1219	93.8	232	5 Q9GZ28	Q9gz28 anemonia su

2	1186	91.2	232	5 Q9GP15	Q9gp15 anemonia su
3	781.5	60.1	227	5 Q9SW86	Q9sw86 condylactis
4	779.5	60.0	227	5 Q9SW11	Q9sw11 condylactis
5	763.5	58.7	227	5 Q9SW85	Q9sw85 radianthus
6	747.5	57.5	227	5 Q8MU45	Q8mu45 condylactis
7	739.5	56.9	231	5 Q8ISF8	Q8isf8 parasicyoni
8	637	49.0	228	5 Q9GP16	Q9gp16 anemonia su
9	561.5	43.2	225	5 Q7ZOW9	Q7zow9 montastraea
10	560.5	43.1	229	5 Q8TSE7	Q8tse7 condylactis
11	559	43.0	224	5 Q8MU48	Q8mu48 montastraea
12	559	43.0	225	5 Q7ZOW5	Q7zow5 montastraea
13	558.5	43.0	227	5 Q9SVT0	Q9svt0 montastraea
14	553.5	42.6	225	5 Q8T5F1	Q8t5f1 montastraea
15	553.5	42.6	236	5 Q8TGU0	Q8tgu0 dendronephr
16	553	42.5	225	5 Q9SUA7	Q9sua7 montastraea
17	551.5	42.4	221	5 Q9SP04	Q9sp04 gonopora t
18	551.5	42.4	227	5 Q962P9	Q962p9 montastraea
19	551.5	42.4	227	5 Q7ZOW8	Q7zow8 montastraea
20	550.5	42.3	225	5 Q7ZOW4	Q7zow4 montastraea
21	548.5	42.2	225	5 Q963F5	Q963f5 montastraea
22	547.5	42.1	227	5 Q7ZOW6	Q7zow6 montastraea
23	545.5	42.0	225	5 Q9U6V8	Q9u6v8 discosoma s
24	543	41.8	225	5 Q8T8T9	Q8t8t9 radianthus
25	541.5	41.7	234	5 Q8T5F2	Q8t5f2 montastraea
26	541.5	41.7	234	5 Q7ZOW7	Q7zow7 montastraea
27	541	41.6	230	5 Q9GTJ7	Q9gtj7 discosoma s
28	536	41.2	231	5 Q8T5E8	Q8t5e8 ricordea fl
29	536	41.2	232	5 Q9U6V7	Q9u6v7 discosoma s
30	535	41.2	228	5 Q86LV4	Q86lv4 radianthus
31	529.5	40.7	234	5 Q8MU47	Q8mu47 montastraea
32	529.5	40.7	266	5 Q9U6Y3	Q9u6y3 discosoma s
33	528	40.6	231	5 Q8T6T8	Q8t6t8 ricordea fl
34	523	40.2	231	5 Q8MU46	Q8mu46 ricordea fl
35	519	39.9	225	5 Q8I6J8	Q8i6j8 trachyphyl
36	513.5	39.5	231	5 Q8T5E9	Q8t5e9 ricordea fl
37	470	36.2	225	5 Q8T5F0	Q8t5f0 scolymia cu
38	469.5	36.1	229	5 Q9U6Y6	Q9u6y6 anemonia ma
39	464	35.7	234	5 Q8T5F3	Q8t5f3 scolymia cu
40	460	35.4	259	5 Q8MMA2	Q8mma2 agaricia fr
41	453.5	34.9	238	5 Q9BLZ0	Q9blz0 ptilosarcus
42	445.5	34.3	238	5 Q9BLV9	Q9blv9 penillia mue
43	434	33.4	239	5 Q8MVA1	Q8mva1 agaricia ag
44	423.5	32.6	214	5 Q86LV7	Q86lv7 meandrina m
45	422.5	32.5	214	5 Q86LV8	Q86lv8 meandrina m

ALIGNMENTS

Madgata

RESULT 1
Q9GZ28 PRELIMINARY; PRT; 232 AA.

AC Q9GZ28; (1)
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE GFP-like chromoprotein FP595.
OS Anemonia sulcata (Snake-locks sea anemone).
OC Eukaryota; Metazoa; ~~Chnidaria~~; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthaeae; Actiniidae; Anemonia.
OX NCBI_TaxID=6108;
RN (1)
RP SEQUENCE FROM N.A. PubMed=10852900;
RX MEDLINE=24408921; PubMed=10852900;
RA Lukyanov K.A., Fradkov A.F., Gurskaya N.G., Matz M.V., Labas Y.A.,
RA Savitsky A.P., Markelov M.L., Zaraisky A.G., Zhao X., Fang Y., Tan W.,
RA Lukyanov S.A.;
RT "Natural Amino Coloration Can Be Determined by a Nonfluorescent Green
RT Fluorescent Protein Homolog."
RL J. Biol. Chem. 275:25879-25882(2000).
DR EMBL; AF246709; AAG02385.1; ...
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR000786; Green_fl_protein.

DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFLUORESCENT.
 DR ProDom; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 232 AA; 25918 MW; CDFE992006F4975E CRC64;

Alignment Scores:
 Pred. No.: 3,41e-73 Length: 232
 Score: 1219.00 Matches: 221
 Percent Similarity: 97.84% Conservativity: 5
 Best Local Similarity: 95.67% Mismatches: 5
 Query Match: 93.77% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-21 (1-707) x Q9GZ28 (1-232)

QY 7 GCCTCCCTGCTGACGAGACCATGCTTCAGGACCACTCGAGGCGACCGTGAACGGC 66
 DB 2 AlaserPheLeuLysLysThrMetProPheLysThrThrLeuGluGlyThrValAsnGly 21

QY 67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCGCAACCCCTCGAGGCGACCCAGGAGATG 126
 DB 22 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41

QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTGCTCCCTTCACATCTCCATCTGTCACCTCC 186
 DB 42 LysileGluValIleGluGlyGlyProLeuProPheAlaPheHisLeuSerThrSer 61

QY 187 TGCATGTACGGCTCCAAAGGCGCTTCATCAAGTACGTGCTCCGCGATCCCGGACTACTTCAAG 246
 DB 62 CysMetTyrGlySerLysThrPheLeuLysTyrValSerGlyLeuProAspTyrPheLys 81

QY 247 CAGTCCCTCCCGAGGCGCTTCACTGGGAGGCGACACCACTACGAGGAGCGCGCTTC 306
 DB 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrThrThrThrThrThrThrThr 101

QY 307 CTGACCGCCACACGAGGCGGCGCCCGTGTATGATGAGAACAAAGCGCGCGCTCGGAG 426
 DB 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGlyArgTyrPhe 141

QY 427 CAGTCCCGCGCGGCGCTGCTACAAAGTACGAGGCGCGCGCTGCTGCTGCTGCTGCTGCT 486
 DB 142 ProAlaThrGluLeuValGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161

QY 487 CTGAGTGCCTCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 546
 DB 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisSerThrThrThrThrThr 181

QY 547 AAGCCCGCTCCCGCTGAGATGCGCGGCTTCCATCTGAGGACCAACCGCATCGAGATC 606
 DB 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGlyAspHisArgIleGluIle 201

QY 607 CTGAGGAGGTGGAGAGGCAAGTGTCTAACAGTACGAGGCGCGCGCTGCTGCTGCTGCT 666
 DB 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221

QY 667 TGGAGCGCGCGCTCCAAAGTGGGCGCAAC 699
 DB 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232

RESULT 2
 Q9GP15
 ID Q9GP15 PRELIMINARY; PRT; 232 AA.
 AC Q9GP15
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Nonfluorescent red protein asCp562.
 OS Anemonia sulcata (Snake-locks sea anemone).
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;

OC Nymphaeae; Actiniidae; Anemonia.
 OX NCBI_TaxID=6108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20570437; PubMed=11121018;
 RA Wiedenmann J., Eike C., Spindler K.D., Funke W.;
 RT "Cracks in the beta -can: fluorescent proteins from Anemonia sulcata
 (Anthozoa, Actinaria).";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:14091-14096 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Wiedenmann J., Eike C., Spindler K.-D., Funke W., Sundin B., Jach G.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF322222; AAG41206.2; -
 DR HSSP; P42212; 1GFL.
 DR GO; GO:0006091; P:energy pathways; IEA.
 DR InterPro; IPR009017; GFP-like.
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFLUORESCENT.
 DR ProDom; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 232 AA; 25956 MW; D7B596FE548097F3 CRC64;

Alignment Scores:
 Pred. No.: 5.37e-71 Length: 232
 Score: 1186.00 Matches: 215
 Percent Similarity: 96.54% Conservativity: 8
 Best Local Similarity: 93.07% Mismatches: 8
 Query Match: 91.23% Indels: 0
 DB: 5 Gaps: 0

US-10-081-864-21 (1-707) x Q9GP15 (1-232)

QY 7 GCCTCCCTGCTGACGAGACCATGCTTCAGGACCACTCGAGGCGACCGTGAACGGC 66
 DB 2 AlaserPheLeuLysLysThrMetProPheLysThrThrLeuGluGlyThrValAsnGly 21

QY 67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCGCAACCCCTCGAGGCGACCCAGGAGATG 126
 DB 22 HistyrPheLysCysThrGlyLysGlyGluGlyAsnProPheGluGlyThrGlnGluMet 41

QY 127 AAGATCGAGGTGATCGAGGCGGCGCCCTGCTCCCTTCACATCTCCATCTGTCACCTCC 186
 DB 42 LysileGluValIleGluGlyGlyProLeuProPheAlaPheHisLeuSerThrSer 61

QY 187 TGCATGTACGGCTCCAAAGGCGCTTCATCAAGTACGTGCTCCGCGATCCCGGACTACTTCAAG 246
 DB 62 CysMetTyrGlySerLysThrPheLeuLysTyrValSerGlyLeuProAspTyrPheLys 81

QY 247 CAGTCCCTCCCGAGGCGCTTCACTGGGAGGCGACACCACTACGAGGAGCGCGCTTC 306
 DB 82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrThrThrThrThrThrThrThr 101

QY 307 CTGACCGCCACACGAGGCGGCGCCCGTGTATGATGAGAACAAAGCGCGCGCTCGGAG 366
 DB 102 LeuThrAlaHisGlnAspThrSerLeuAspGlyAspCysLeuValTyrLysValLysIle 121

QY 367 CTGGGCGCAACACTTCCCGCGAGGCGGCGCCCGTGTATGATGAGAACAAAGCGCGCGCT 426
 DB 122 LeuGlyAsnAsnPheProAlaAspGlyProValMetGlnAsnLysAlaGluArgTyrPhe 141

QY 427 CCCTCCACCGAGATCGGTGTACGAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
 DB 142 ProAlaThrGluLeuValGluValAspGlyValLeuArgGlyGlnSerLeuMetAla 161

QY 487 CTGAGTGCCTCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 546
 DB 162 LeuLysCysProGlyGlyArgHisLeuThrCysHisLeuHisSerThrThrThrThrThr 181

QY 547 AAGCCCGCTCCCGCTGAGATGCGCGGCTTCCATCTGAGGACCAACCGCATCGAGATC 606
 DB 182 LysProAlaSerAlaLeuLysMetProGlyPheHisPheGlyAspHisArgIleGluIle 201

QY 607 CTGAGGAGGTGGAGAGGCAAGTGTCTAACAGTACGAGGCGCGCGCTGCTGCTGCTGCT 666
 DB 202 MetGluGluValGluLysGlyLysCysTyrLysGlnTyrGluAlaValGlyArgTyr 221

QY 667 TGGAGCGCGCGCTCCAAAGTGGGCGCAAC 699
 DB 222 CysAspAlaAlaProSerLysLeuGlyHisAsn 232


```

Db      62 CysAlaTyrGlySerLysThrPheLeuLeuHisThrSerGlyLeuLeuProAspTyrPheLys 81
QY      247 CAGTCCCTCCCGAGGCGTTACCTGGAGCGCGACCACTACGAGGACGGCGGCTTC 306
Db      82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrLeuTyrGluAspGlyGlyVal 101
QY      307 CTGACCGCCACCCAGGACACCTCCCTGAGCGGCGACTGCCTGGTGTACAAAGTGAAGTC 366
Db      102 LeuThrAlaHisGlnAspThrSerLeuGluGlyAsnCysLeuAsnTyrLysValLysVal 121
QY      367 CTGGCAACAACCTCCCGCGAGCGCGCGCTGTATGAGAACAAAGCGCGCGCGCTGGAG 426
Db      122 LeuGlyThrAsnProAlaAspGlyProValMetLysAsnLeuSerGlyGlyTyrGlu 141
QY      427 CCCTCCACCGAGATCGTGTACAGGTGTACAGCGCGCTGTGGCGCGCGAGTCCCTGATGCC 486
Db      142 ProCysThrGluLeuValTyrGlnAspAsnGlyValLeuArgGlyArgAsnValMetAla 161
QY      487 CTGAGTGCCTCCCGCGGTGCGCACCTGACCTGCACCTGCACACCACTACCGCTCCCAAG 546
Db      162 LeuLysValSerGlyArgProProLeuLeuLeuCysHisLeuHisSerThrTyrArgSerLys 181
QY      547 AAGCCCGCTCCCGCTTCAAGATCGCGGCTTCACCTTCGAGGACCAACCGATCGAGATC 606
Db      182 Lys---AlaCysAlaLeuThrMetProGlyPheHisPheAlaAspLeuArgLeuGln--- 199
QY      607 CTGAGGAGGTGGAGAGGCAAGTGTACAAAGCAGTACAGGCGCGCGCGTGGCGCGTAC 666
Db      200 MetProLysLysLysLysAspGlyTyrPheGluLeuTyrGluAlaSerValAlaArgTyr 219
QY      667 TGCAGCGCGCGCGCTCCCAAG 687
Db      220 SerAsp---LeuProGluLys 225

RESULT 5
Q95W85
ID Q95W85 PRELIMINARY; PRT; 227 AA.
AC Q95W85;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Radianthus macrodactylus (Sea anemone) (Heteractis macrodactylus).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Stichodactylidae; Heteractis.
OX NCBI_TaxID=175771;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21538626; PubMed=11692051;
RA Gurskaya N.G.; Fradkov A.F.; Tersikh A.; Matz M.V.; Labas Y.A.;
RA Martynov V.I.; Yanushevich Y.G.; Lukyanov K.A.; Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
RT proteins(1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF363776; AAL27538.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 227 AA; 25637 MW; CB40899E95E7EC64 CRC64;

Alignment Scores:
Pred. No.: 7.16e-43 Length: 227
Score: 763.50 Matches: 141
Percent Similarity: 77.09% Conservative: 34
Best Local Similarity: 62.11% Mismatches: 49
Query Match: 58.73% Indels: 3
DB: 5 Gaps: 3

US-10-081-864-21 (1-707) x Q95W85 (1-227)

```

```

QY      7 GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGCG 66
Db      2 AlaGlyLeuLeuLysGluSerMetArgLeuLysMetTyrMetGluGlyThrValAsnGly 21
QY      67 CACTACTTCAAGTGCACCGGCAAGGCGAGGCAACCCCTCGAGGCGACCCAGGAGATG 126
Db      22 HisTyrPheLysCysGluGlyGluGlyAspGlyAsnProPheThrGlyThrGlnSerMet 41
QY      127 AAGATCGAGGTATCGAGGCGGCGCGCTTCGCTTCGCTCCACATCCTGTCACCTCC 186
Db      42 ArgIleHisValThrGluGlyAlaProLeuProPheAlaPheAspIleLeuAlaProCys 61
QY      187 TGCATGTACGGTCCCAAGCGCTTCATCAAGTACGTGTCGGCATCCCGCATCTACTTCAAG 246
Db      62 CysGluTyrGlySerArgThrPheValHisThrAlaGluLeuProAspPhePheLys 81
QY      247 CAGTCCCTCCCGAGGCTTACCTGGAGCGCGACCACTACGAGGACGGCGGCTTC 306
Db      82 GlnSerPheProGluGlyPheThrTrpGluArgThrThrTyrGluAspGlyGlyTyr 101
QY      307 CTGACCGCGCCACCGAGCACCTCCCTGAGCGCGAGTCCCTGGTGTACAAAGTGAAGTC 366
Db      102 LeuThrAlaHisGlnAspThrSerLeuGluGlyAsnCysLeuLeuTyrLysValLysVal 121
QY      367 CTGGCAACAACCTCCCGCGAGCGCGCGCTGTATGAGAACAAAGCGCGCGCGTGGAG 426
Db      122 LeuGlyThrAsnProAlaAspGlyProValMetLysAsnLysSerGlyGlyTyrGlu 141
QY      427 CCCTCCACCGAGATCGTGTACAGGTGTACAGCGCGTGTGGCGCGCGAGTCCCTGATGCC 486
Db      142 ProCysThrGluValTyrProGluAsnGlyValLeuCysGlyArgAsnValMetAla 161
QY      487 CTGAGTGCCTCCCGCGGTGCGCACCTGACCTGCACCTGCACACCACTACCGCTCCCAAG 546
Db      162 LeuLysVal---GlyAspArgArgLeuLeuLeuCysHisLeuTyrThrSerTyrArgSerLys 180
QY      547 AAGCCCGCTCCCGCTTCAAGATCGCGGCTTCACCTTCGAGGACCAACCGATCGAGATC 606
Db      181 LysAlaValArgAlaLeuThrMetProGlyPheHisPheThrAspIleArgLeuGln--- 199
QY      607 CTGAGGAGGTGGAGAGGCAAGTGTACAAAGCAGTACAGGCGCGCGCGTGGCGGCTAC 666
Db      200 MetProArgLysLysLysLysAspGlyTyrPheGluLeuTyrGluAlaSerValAlaArgTyr 219
QY      667 TGCAGCGCGCGCGCTCCCAAG 687
Db      220 SerAsp---LeuProGluLys 225

RESULT 6
Q95W85
ID Q95W85 PRELIMINARY; PRT; 227 AA.
AC Q95W85;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein-like protein.
OS Condylactis gigantea (giant anemone).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynantheae; Actinidae; Condylactis.
OX NCBI_TaxID=47073;
RN [1]
RP SEQUENCE FROM N.A.
RX Matz M.V.; Lukyanov S.A.;
RT "Diversity and evolution of GFP-like fluorescent proteins.";
RT Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037777; AAK71343.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 227 AA; 25384 MW; D3C6B02F490F3D21 CRC64;

```


Q9GP16
ID Q9GP16 PRELIMINARY; PRT; 228 AA.
AC Q9GP16;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Green fluorescent protein asFP499 (Green fluorescent protein as(s)FP499).
OS Anemonia sulcata (Snake-locks sea anemone).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nyantheae; Actinidae; Anemonia.
OX NCBI_TaxID=6108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20570437; PubMed=11121018;
RA Wiedenmann J., Elke C., Spindler K.-D., Funke W.;
RT "Cracks in the beta-can: Fluorescent proteins from Anemonia sulcata (Anthozoa, Actiniaria).";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14091-14096(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Leutenegger A., Wiedenmann J.;
RT "Photobiology of Anemonia sulcata.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF322221; AAG41205.1; -;
DR EMBL; AF545827; AAN52735.1; -;
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl protein; 1.
SQ SEQUENCE 228 AA; 25369 MW; 89CEBF7B1ECDE73 CRC64;

Alignment Scores:
Pred. No.: 1.89e-34 Length: 228
Score: 637.00 Matches: 126
Percent Similarity: 67.54% Conservative: 28
Best Local Similarity: 55.26% Mismatches: 70
Query Match: 49.00% Indels: 4
DB: 5 Gaps: 2

US-10-081-864-21 (1-707) x Q9GP16 (1-228)

QY 16 CTGACCGAGACCATGCCCTTCAGACACCATCGAGGACCGTGAAGCCGCTACTTC 75
Db 5 IleYsGluThrMetArgValGlnLeuSerMetGluGlySerValAsnTyrHisAlaPhe 24
QY 76 AAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGACCCAGGAGATGAAGATCGAG 135
Db 25 LysCysThrGlyLysGlyGlyLysProTyrGluGlyThrGlnSerLeuAsnIleThr 44
QY 136 GTGATCAGAGGGGCGCCCTTGCCTTCGCTTCACATCCTGTCACATCCTGTCATGTAC 195
Db 45 IleThrGluGlyGlyProLeuProPheAlaPheAspIleLeuSerHisAlaPheGlnTyr 64
QY 196 GGCTCCAAAGCCCTTCATAGTACGTGTCGGGATCCCGGATCTTCAAGCAGTCCCTC 255
Db 65 GlyIleYsValPheAlaLysTyrProLysGluIleProAspPheLysGlnSerLeu 84
QY 256 CCGAGGGCTTCACCTGGGAGCGCACCCACCATCAGGAGCGCGGCTTCCTGACCGCC 315
Db 85 ProGlyGlyPheSerTrpGluArgValSerThrTyrGluAspGlyGlyValLeuSerAla 104
QY 316 CACCAGACACCTCCCTGGACGGGACTGCTGCTGCTACAGGTGAAGATCCTGGGCAAC 375
Db 105 ThrGlnGluThrSerLeuGlnGlyAspCysIleCysLysValLysValLeuGlyThr 124
QY 376 AACTTCCCGCGGACCGCCGCTGATGCAGAACAGGCGCGGCTGGGAGCCCTCCACC 435
Db 125 AsnPheProAlaAsnGlyProValMetGlnLysThrCysGlyTrpGluProSerThr 144
QY 436 GAGATCGTACGAGGTGACGCGGTGTCGGCGGCGACCTCCCTGATGCGCTGAGGTGC 495

Db 145 GluThrValIleProArgAspGlyGlyLeuLeuArgAspThrProAlaLeuMetLeu 164
QY 496 CCGGGGGTGGCCACCTGACCTGCCACCTGCACACCATCAGCTCCCAAGAACGCCGCC 555
Db 165 AlaAspGlyGlyHisLeuSerCysPheMetGluThrThrTyrLysSerLysLys----- 182
QY 556 TCGGCGCTGAAGATGCCGCTTCCTTCGAGGACCCCGCATCGAGATCCTGGAGGAG 615
Db 183 ---GluValLysLeuProGluLeuHisPheHisLeuArgMetGluLysLeuAsnIle 201
QY 616 GTGGAGAGGCGCAAGTCTACAGGAGTACGAGGCGCGCTGCGCGCTTACTGCGACGCC 675
Db 202 SerAspAspTrpLysThrValGluGlnHisGluSerValAlaSerTyr---SerGln 220
QY 676 GCCCCTCCAAAGCTGGGCCAACAC 699
Db 221 ValProSerLysLeuGlyHisAsn 228
RESULT 9
Q720W9
ID Q720W9 PRELIMINARY; PRT; 225 AA.
AC Q720W9;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Red fluorescent protein.
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mc1.
RX MEDLINE=22689801; PubMed=12777529;
RA Kelmanson I.V., Matz M.V.;
RT "Molecular Basis and Evolutionary Origins of Color Diversity in Great Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL Mol. Biol. Evol. 20:1125-1133(2003).
DR EMBL; AY181552; AAO61598.1; -;
SQ SEQUENCE 225 AA; 25822 MW; 830871F6936953E2 CRC64;

Alignment Scores:
Pred. No.: 2e-29 Length: 225
Score: 561.50 Matches: 109
Percent Similarity: 63.64% Conservative: 31
Best Local Similarity: 49.55% Mismatches: 75
Query Match: 43.19% Indels: 5
DB: 5 Gaps: 2

US-10-081-864-21 (1-707) x Q720W9 (1-225)

QY 10 TCCCTGTGACCGAGACCATGCCCTTCAGGACCATCGAGGACCGTGAAGCGCCAC 69
Db 2 SerValIleYsSerValMetLysLysLeuArgMetGluGlySerValAsnGlyHis 21
QY 70 TACTTCAAGTGCACCGGCAAGGCGGAGGCAACCCCTCGAGGACCCAGGAGATGAAG 129
Db 22 AsnPheValIleValGlyGluGlyLysProTyrGluGlyThrGlnSerMetAsp 41
QY 130 ATCAGAGTGTACGAGGCGCGCCCTGCTTCGCTTCACATCCTGCTCCACATCCTCTG 189
Db 42 LeuThrValLysGluGlyAlaProLeuProPheAlaTyrAspIleMetThrThrValPhe 61
QY 190 ATGTACGGCTCCAAAGCCTTCATCAAGTACGTGTCCGGCATCCCGGACTACTTCAAGCAG 249
Db 62 HisTyrGlyAsnArgValPheAlaLysTyrProLysHisIleProAspTyrPheLysGln 81
QY 250 TCCCTCCCGGCGCTTCACCTGGGAGCGGACACCATCAGGACCGCGGCTTCCTG 309
Db 82 MetPheProGluGlyTyrSerTrpGluArgSerMetAsnPheGluAspGlyIleCys 101
QY 310 ACCGCCACGAGGACACCTCCCTGAGCGGCGACTGCTGTGTACAGGTGAAGATCCTG 369

```

Db      102 ThrAlaArgAsnGluLeuThrMetGluGlyAspCysPheAsnLysValArgPheAsp 121
QY      370 GGCAACAACATCCCGCGGAGCGCCCGTGTATGAGAAACAAGCGCGCGCTGGAGCCC 429
Db      122 GlyValAsnPheProAsnGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
QY      430 TCACACGAGATCGTGTACAGAGGTGGAGCGGTGCTGGCGGCGCAGTCCCTGTAGTGGCCCTG 489
Db      142 SerThrGluLysMetTrpValArgAspGlyValLeuThrGlyAspIleAsnMetAlaLeu 161
QY      490 GAGTGGCCCGCGGTGCGCCACCTGACCTGCACCTGCACACCACTACCGCTCCAGAAG 549
Db      162 LeuLeuGluGlyGlyHisTyrArgCysAspPheArgThrThrArgAlaLysLys 181
QY      550 CCGCCCTCCCGCTGAGATGCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATCCCTG 609
Db      182 -----LysGlyValLysLeuProAspTyrHisPheValAspHisSerIleGluIleLeu 199
QY      610 -----GAGGAGGTGGAGAGGCAAGTGTACAGCAGTACGAGGCGCGCTGGGC 660
Db      200 ArgHisAspLysGluTyrThrGluValLysLeuTyrGluHisAlaGluAlaHisSerGly 219

RESULT 10
Q8T5E7 ID Q8T5E7 PRELIMINARY; PRT; 229 AA.
AC Q8T5E7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CgIFP-g.
OS Condylactis gigantea (Giant anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Acriniaria;
OC Nyanthesea; Actinidae; Condylactis.
OX NCBI_TaxID=47073;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21927629; PubMed=11929936;
RA Labas V.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F.,
RA Lukyanov K.A., Lukyanov S.A., Matz M.V.;
RT "Diversity and evolution of the green fluorescent protein family.";
DR GO; GO:0006091; P:energy pathways; IEA.
DR EMBL; AY037776; AAK71342.1; -.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 229 AA; 25731 MW; A5E4E1CCFE8E85B4 CRC64;

Alignment Scores:
Pred. No.: 2,34e-29 Length: 229
Score: 560.50 Matches: 113
Percent Similarity: 60.79% Conservative: 25
Best Local Similarity: 49.78% Mismatches: 84
Query Match: 43.12% Indels: 5
DB: Gaps: 5

US-10-081-864-21 (1-707) x Q8T5E7 (1-229)
QY 16 CTGACGAGACCATGCCCTTCAGGACCACTGAGGGGACCGTGAACGGCCACTACTTC 75
Db 5 IleLysGluThrMetArgSerLysValTyrMetGluGlyAspValAsnAsnHisAlaPhe 24
QY 76 AAGTGACCGCGCAAGGCGGAGGCAACCCCTCGAGGACCCAGGAGATGAGATCGAG 135
Db 25 LysCysThrAlaValGlyGluGlyLysProTyrLysGlySerGlnAspLeuThrIleThr 44
QY 136 GTGATCAGGCGGCGCCCTGCGCTTCGCTTCCATCCTGTCCACCTCCCTCGATGAC 195
Db 45 ValThrGluGlyGlyProLeuProPheAlaPheAspIleLeuSerHisAlaPheGlnTyr 64
QY 196 GGCTCCCAAGGCTTCATCAAGTAGTCGTCCGGCATCCCGACTACTTCAGCAGTCCCTC 255

```

```

Db      65 GlyAsnLysValPheThrAspTyrProAspAspIleProAspPhePheLysGlnSerLeu 84
QY      256 CCCGAGGGCTTCACCTGGGAGCGCACCACTACGAGGAGCGGGCTTCTGTGACCGCC 315
Db      85 SerAspGlyPheThrTrpArgValSer***Tyr***GlyGlyValLeuThrVal 104
QY      316 CACGAGGACACCTCCCTGGAGCGGACTGCTGTGTACAGGTGAAGTGAAGTCTGGGCAAC 375
Db      105 ThrGlnAspThrSerLeuLysGlyAspCysIleLeuCysAsnIleLysValHisGlyThr 124
QY      376 AACTTCCCGCGGAGCGCGCTGATGAGAACACAGGCGCGCGCTGGAGGCC---TCC 432
Db      125 AsnPheProGluAsnGlyProValMetGlnAsnLysThrAspGlyTrpGluProSerSer 144
QY      433 ACCGAGATCGTGTACGAGGTGACGCGCTGCTGCGCGGCCACATCCCTGATGCGCCCTGGAG 492
Db      145 ThrGluThrValIleProGlnAspGlyGlyIleValAlaAlaArgSerProAlaLeuArg 164
QY      493 TGCCCGCGGCGTGGCGCACCTGACCTGCCACCTGACAGCGCGCGCTCCAGAGAGGCC 552
Db      165 LeuArgAspLysGlyHisLeuIleCysHisMetGluThrThrTyr-----LysPro 181
QY      553 GCTCCGCGCTGAGATGCCCGGCTTCCACTTCGAGGACCAACCGCATCCGATCCCTGGAG 612
Db      182 AsnLysGluValLysLeuProGluLeuHisPheHisLeuArgMetGluLysLeuSer 201
QY      613 GAGGTGGAGAGGCAAGTGTCTACAGCAGTACAGAGCGCGCGCTGCGGCGCTACTGCGAC 672
Db      202 ValSerAspAspGlyLysThrIleLysGlnHisGluTyrValValAlaSerTyr---Ser 220
QY      673 GCGCGCGCTCCCAAGCTGGC 693
Db      221 LysValProSerLysIleGly 227

RESULT 11
Q8MU48 ID Q8MU48 PRELIMINARY; PRT; 224 AA.
AC Q8MU48;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein-like protein.
OS Montastraea annularis (boulder star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=48500;
RN [1]
RP SEQUENCE FROM N.A.
RA Matz M.V., Lukyanov S.A.;
RT "Diversity and evolution of GFP-like fluorescent proteins.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037766; AAK71332.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 224 AA; 25452 MW; 1C4D85F87D6315EC CRC64;

Alignment Scores:
Pred. No.: 2,94e-29 Length: 224
Score: 559.00 Matches: 108
Percent Similarity: 63.18% Conservative: 31
Best Local Similarity: 49.09% Mismatches: 75
Query Match: 43.00% Indels: 6
DB: Gaps: 5

US-10-081-864-21 (1-707) x Q8MU48 (1-224)
QY 10 TCCTCGCTGACGAGACCATGCCCTTCAGGACCACTGAGGCGCGTGAACGGCCAC 69
Db 2 SerMetIleLysProGluMetLysIleLysMetArgMetAspGlyAlaValAsnGlyHis 21

```



```

SQ SEQUENCE 227 AA; 26035 MW; 2C41B38DCB84279C CRC64;
Alignment Scores:
Pred. No.: 3,17e-29 Length: 227
Score: 558.50 Matches: 112
Percent Similarity: 63.68% Conservatives: 30
Best Local Similarity: 50.22% Mismatches: 72
Query Match: 42.96% Indels: 9
DB: 5 Gaps: 4

US-10-081-864-21 (1-707) x Q95VT0 (1-227)
QY 10 TCCTCTGCTGACGAGACCATGCGCTTCAGGACACCATCGAGGCGACCGTGAACGGCCAC 69
Db 2 SerValIleLysProAspMetLysLysLeuArgMetGluGlyAlaValAsnGlyHis 21
QY 70 TACTTCAAGTGCACCGGAGGCGCAACCCCTCGAGGCGACCCAGGAGATGAAG 129
Db 22 LysPheValIleGluGlyAspGlyLysGlyLysProPheGluGlyThrGlnSerMetAsp 41
QY 130 ATCAGAGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCTGCTCCACCTCTGC 189
Db 42 LeuThrValIleGluGlyAlaProLeuProPheAlaTyrAspIleLeuThrValPhe 61
QY 190 ATGTAGGCTCCAGGCTTCATCAAGTACGTCTCGGATCCCGACTCCGACTCTCAAGCAG 249
Db 62 AspTyrGlyAsnArgValPheAlaLysTyrProGlnAspIleThrAspTyrPheLysGln 81
QY 250 TCCTCTCGGAGGCTTCACCTGGAGCGCACCCACCTACGAGGCGCGGCTTCCTG 309
Db 82 ThrPheProGluGlyTyrPheTyrPgluArgSerMetThrTyrGluAspGlnGlyLysCys 101
QY 310 ACCGCCACACAGAC---ACCTCCCTGGAGCGC-----GACTGCTGTGTACAGGTG 360
Db 102 IleAlaThrAsnAspIleThrMetMetLysGlyValAspAspCysPheValTyrLysIle 121
QY 361 AAGATCTGGGCAACAACTTCCCGCGGACCGGCTGATGAGAACAGGCGCGCGC 420
Db 122 ArgPheAspGlyValAsnPheProAlaAsnGlyProValMetGlnArgLysThrLeuLys 141
QY 421 TGGAGCGCTTCCACGAGATCGTGTACAGGTGGACGGGTGTGTGCGGCGCCAGTCCCTG 480
Db 142 TrpGluProSerThrGluIleMetTyrAlaArgAspGlyValLeuLysGlyAspValAsn 161
QY 481 ATGGCTGTGAGTGGCGCGGCTGCGACCTGACCTGCGACCTGCGACACCTACCGC 540
Db 162 MetAlaLeuLeuLeuGluGlyGlyGlyHisTyrArgCysAspPheLysThrThrTyrLys 181
QY 541 TCCAAGAAGCGCGCTCCGCGCTCGAGATCGCGGCTTCCACTTCGAGGACCAACCGCATC 600
Db 182 AlaLysLys-----ValValArgLeuProAspTyrHisPheValAspHisArgIle 198
QY 601 GAGATCTCTG-----GAGGAGGTGGAGAGGCGAGTGTCTACAGCAGTACGAGGCG 651
Db 199 GluIleValSerHisAspLysAspTyrAsnLysValLysLeuTyrGluHisAlaGluAla 218
QY 652 GCGTGGCG 660
Db 219 *****Gly 221

RESULT 14
Q8T5F1 PRELIMINARY; PRT; 225 AA.
AC Q8T5F1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE McavFP. 7.5.
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastraea.
ON NCB1_TaxID=63558;
RX SEQUENCE FROM N.A.
RX MEDLINE=21927629; PubMed=11929996;
RA Labas Y.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F.,
RA Lukyanov K.A., Lukyanov S.A., Matz M.V.;
RT "Diversity and evolution of the green fluorescent protein family.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
DR EMBL; AY037770; AAK71336.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR PRODOM; PD013756; Green fl protein; 1.
SQ SEQUENCE 225 AA; 25866 MW; 820C89437F8BDB32 CRC64;
Alignment Scores:
Pred. No.: 6.83e-29 Length: 225
Score: 553.50 Matches: 108
Percent Similarity: 63.18% Conservatives: 31
Best Local Similarity: 49.09% Mismatches: 76
Query Match: 42.58% Indels: 5
DB: 5 Gaps: 2

US-10-081-864-21 (1-707) x Q8T5F1 (1-225)
QY 10 TCCTCTGCTGACGAGACCATGCGCTTCAGGACACCATCGAGGCGACCGTGAACGGCCAC 69
Db 2 SerValIleLysSerValMetLysLysLeuArgMetGluGlySerValAsnGlyHis 21
QY 70 TACTTCAAGTGCACCGGAGGCGCAACCCCTCGAGGCGACCCAGGAGATGAAG 129
Db 22 AsnPheValIleValGlyGluGlyLysProTyrGluGlyThrGlnSerMetAsp 41
QY 130 ATCAGAGTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCTGCTCCACCTCTGC 189
Db 42 LeuThrValLysGluGlyAlaProLeuProPheAlaTyrAspIleMetThrThrValPhe 61
QY 190 ATGTAGGCTCCAGGCTTCATCAAGTACGTCTCGGATCCCGACTCCCGACTCTCAAGCAG 249
Db 62 HisTyrGlyAsnArgValPheAlaLysTyrProLysHisIleProAspTyrPheLysGln 81
QY 250 TCCTCTCGGAGGCTTCACCTGGAGCGCACCCACCTACGAGGCGCGGCTTCCTG 309
Db 82 MetPheProGluGlyTyrSerTrpGluArgSerMetAsnPheGluGlyGlyLysCys 101
QY 310 ACCGCCACACGAGCACCTCCCTGAGCGGCGACTGCTGTGTACAGGTGAAGATCTCTG 369
Db 102 ThrAlaArgAsnGluIleThrMetGluGlyAspCysPhePheAsnLysValArgPheAsp 121
QY 370 GGCACAACTTCCCGCGCGACGCGCTGTGCGCGCGCGCTGCTGCGCGCGCTGCTG 489
Db 122 GlyValAsnPheProAsnGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
QY 430 TCCACGAGATCGTGTACAGGTGAGCGGCTGTGCGCGCGCGCTGCTGCGCGCGCTG 489
Db 142 SerThrGluLysMetTyrValArgAspGlyValLeuThrGlyAspIleAsnMetAlaLeu 161
QY 490 GAGTGGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCG 549
Db 162 LeuLeuGluGlyGlyGlyHisTyrArgCysAspPheArgThrThrTyrArgAlaLysLys 181
QY 550 CCGCGCTCGCGCTGAGATCGCGCTTCACCTTCGAGGCGACCCAGGATCGAGTCTCTG 609
Db 182 -----LysGlyValLysLeuProAspTyrHisPheGluAspHisSerIleGluLeu 199
QY 610 -----GAGGAGGTGGAGAGGCGAGTGTCTACAGCAGTACGAGGCGCGCTGGGC 660
Db 200 ArgHisAspLysGluTyrThrGluValLysLeuTyrGluHisAlaGluAlaHisSerGly 219

RESULT 15
Q8T6U0
ID Q8T6U0
AC Q8T6U0;

```

DT 01-JUN-2002 (TREMELrel. 21, Created)
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Green fluorescent protein.
OS Dendronephthya sp. SSU-2002.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
OC Nephthelidae; Dendronephthya.
OX NCBI TaxID=191210;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21927629; PubMed=11929996;
RA Labas Y.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F.,
RA Lukyanov K.A., Lukyanov S.A., Matz M.V.;
RT "Diversity and evolution of the green fluorescent protein family";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
DR EMBL; AF420591; AAM10625.1; -.
DR GO; GO:0006091; P:energy pathways, IEA.
DR InterPro; IPR009017; GPP-like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GPP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 236 AA; 26840 MW; CE1707CF9334A90 CRC64;

Alignment Scores:

Pred. No.:	6.86e-29	Length:	236
Score:	553.50	Matches:	108
Percent Similarity:	61.67%	Conservative:	32
Best Local Similarity:	47.58%	Mismatches:	84
Query Match:	42.58%	Indels:	3
DB:	5	Gaps:	1

US-10-081-864-21 (1-707) x Q8T6U0 (1-236)

QY	10	TCCTGTGTGACGAGACCATCCCTTCAGGACCACTCGAGGCGACCGTGAACGGCCAC	69
DB	2	AsnLeuIleLysGluAspMetArgValLysValHisMetGluGlyAsnValAsnGlyHis	21
QY	70	TACTTCAAGTCACGGGAGGGGAGGGGACCCCTCGAGGCGACCCAGGAGATGAAG	129
DB	22	AlaPheValIleGluGlyGluGlyGlyArgProTyrGluGlyThrGlnThrLeuAsn	41
QY	130	ATCGAGGTGATCGAGGGGGGGCCCTGCTCCCTTCCTCCACATCCCTGCTCCCTCTGC	189
DB	42	LeuThrValLysGluGlyAlaProLeuProPheSerTyrAspIleLeuThrAlaLeu	61
QY	190	ATGTACGGCTCCAGGGCTTCATCAAGTAGTGTCCGGCATCCCGACTACTTCAAGCAG	249
DB	62	HisTyrGlyAsnArgValPheThrGluTyrProAlaAspIleThrAspTyrPheLysGln	81
QY	250	TCCCTCCCGAGGCTTCACCTGGGAGGCGACACACCTACGAGGCGGCGCTTCTCTG	309
DB	82	SerPheProGluGlyTyrSerTrpGluArgThrMetThrTyrGluAspLysGlyIleCys	101
QY	310	ACCGCCACACGAGACACTCCCTGCGGCGGACTGCTGTGTACAGGTGAAGATCCTG	369
DB	102	ThrIleArgSerAspIleSerLeuGluGlyAspCysPheGlnAsnIleArgPheAsn	121
QY	370	GGCAACACACTCCCGCGGCGGCGCTGTGTGATGATGAGGCGGCGCTGGGAGCCC	429
DB	122	GlyMetAsnPheProProAsnGlyProValMetGlnLysLysThrLeuLysTrpGluPro	141
QY	430	TCCACCGAGATCGGTGACGAGGTGGAGGGCTGTGCGGCGGCGGCTGATGGCCCTG	489
DB	142	SerThrGluLysLeuHisValArgAspGlyLeuLeuValGlyAsnIleAsnMetAlaLeu	161
QY	490	GAGTGCCCGGCGGTCCCGACCTGACCTGCGACCTGACACACCTACCGCTCCAGAG	549
DB	162	LeuLeuGluGlyGlyHisTyrLeuCysAspPheLysThrThrTyrLysAlaLysLys	181
QY	550	CCGCGCTCCGCGCTGAGATGCCGGCTTCCACTTCGAGGACACCGCATCCGAGATCCTG	609
DB	182	-----ValValGlnLeuProAspTyrHisPheValAspHisArgIleGluIleLeu	198

QY	610	GAGGAGGTGGAGAGGGCAAGTGCTACAAGCACTACGAGGCGCGCTGGCGCTACTGC	669
DB	199	SerAsnAspSerAspTyrAsnLysValLysLeuTyrGluHisGlyValalaArgTyrSer	218
QY	670	GACGCGCGCCCCCTCCAAGCTG	690
DB	219	ProLeuProLysSerGlyLeu	225

Search completed: July 29, 2004, 14:36:17
Job time : 45.1137 secs

Pred. No. is the number of results predicted by chance to have a

```
source
1. .707
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="non-aggregating mutant"

ORIGIN
Query Match 100.0%; Score 707; DB 6; Length 707;
Best Local Similarity 100.0%; Pred. No. 6.7e-79;
Matches 707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCGCGCTCCCTGCTGACCGAGACCATGCGCTTCAGGACCAACCATCGAGGGCACCGTG 60
Db 1 GATCGCGCTCCCTGCTGACCGAGACCATGCGCTTCAGGACCAACCATCGAGGGCACCGTG 60

QY 61 AACGGCCACTACTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTCGAGGGCACCCAG 120
Db 61 AACGGCCACTACTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTCGAGGGCACCCAG 120

QY 121 GAGATGAAGATCGAGGTGATCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180
Db 121 GAGATGAAGATCGAGGTGATCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180

QY 181 ACCTCTGCTAGTACGGCTTCAAGGCTTATCAAGTACGTGTCCGGCANTCCCGACTAC 240
Db 181 ACCTCTGCTAGTACGGCTTCAAGGCTTATCAAGTACGTGTCCGGCANTCCCGACTAC 240

QY 241 TTCAAGCAGTCCCTCCCGAGGGGTTTCACTTGGGAGCGGACCAACCTACGAGGAGCGG 300
Db 241 TTCAAGCAGTCCCTCCCGAGGGGTTTCACTTGGGAGCGGACCAACCTACGAGGAGCGG 300

QY 301 GGCTTCTGACCGGCGCACAGGACACTTCCCTGGAAGCGGCACTGCTGTGTCAAGGTG 360
Db 301 GGCTTCTGACCGGCGCACAGGACACTTCCCTGGAAGCGGCACTGCTGTGTCAAGGTG 360

QY 361 AAGATCTGGGCAACACTTCCCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 361 AAGATCTGGGCAACACTTCCCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420

QY 421 TGGAGAGCCCTCCACCGAGATCGTGTACAGGTGAGCGGCGGCGGCGGCGGCGGCGGCG 480
Db 421 TGGAGAGCCCTCCACCGAGATCGTGTACAGGTGAGCGGCGGCGGCGGCGGCGGCGG 480

QY 481 ATGCGCTTGGAGTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 481 ATGCGCTTGGAGTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540

QY 541 TCCGAAGAGCCCGCTCCGCGCTCAAGATGCGCGGCTTCCACTTCGAGGACCAACCGATC 600
Db 541 TCCGAAGAGCCCGCTCCGCGCTCAAGATGCGCGGCTTCCACTTCGAGGACCAACCGATC 600

QY 601 GAGATCTGGAGGAGGTGGAGAGGGCAAGTGTACAAGCAGTACGAGGCGGCGGCGGCGG 660
Db 601 GAGATCTGGAGGAGGTGGAGAGGGCAAGTGTACAAGCAGTACGAGGCGGCGGCGGCGG 660

QY 661 CGCTACTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 707
Db 661 CGCTACTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 707

RESULT 2
AY233273
LOCUS
DEFINITION Synthetic construct kindling fluorescent protein mRNA, complete cds.
ACCESSION AY233273
VERSION AY233273.1
KEYWORDS GI:28629492
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 699)
AUTHORS Chudakov,D.M., Belousov,V.V., Zaraisky,A.G., Novoselov,V.V.,

Staroverov,D.B., Zorov,D.B., Lukyanov,S. and Lukyanov,K.A.
Kindling fluorescent proteins for precise in vivo photolabeling
Nat. Biotechnol. 21 (2), 191-194 (2003)
22447893
PUBMED
12524551
REFERENCE
2 (bases 1 to 699)
AUTHORS Chudakov,D.M., Lukyanov,K.A. and Lukyanov,S.
Direct Submission
Submitted (11-FEB-2003) Institute of Bioorganic Chemistry,
Mickluho-Makaya, Moscow 117997, Russia
Location/Qualifiers
1. .699
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
1. .699
/note="GFP-like protein; capable of photoconversion from nonfluorescent to red fluorescent form; KPPI; KFP-red"
/codon_start=1
/transl_table=1
/product="kindling fluorescent protein"
/protein_id="AAO43187.1"
/db_xref="GI:28629492"
/translation="MASLLTETMFKTTIEGTNGHCFKICGKGEHPFEGTQEMKIE
VIEGPLEFAFHLSTSCMYGSKTFKYVSGIPDYFKQSPFEGTERTTVEYDGL
TAHQDTSIDGDLVYKVLGNFPADQPMQNVGRWEPGTIVYEDVGLRQGLM
ALKCPGGRHLTCHLHTTYRSKPASALKMPGFHEDHRIEIMEVEKCKYQVEAAV
GRYCDAAPSKLGN"
ORIGIN
Query Match 95.0%; Score 672; DB 12; Length 699;
Best Local Similarity 97.8%; Pred. No. 1.5e-74;
Matches 681; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 7 GCCTCCCTGCTGACCGAGACCATGCTTCAGGACCAACCTCGAGGGCACCGTGAACGGC 66
Db 4 GCCTCCCTGCTGACCGAGACCATGCTTCAGGACCAACCTCGAGGGCACCGTGAACGGC 63

QY 67 CACTACTTCAAGTGACCGGCAAGGGCGAGGCGAACCCTCGAGGGCACCGAGAGATG 126
Db 64 CACTGCTTCAAGTGATCGGCAAGGGCGAGGCGAACCCTCGAGGGCACCGAGAGATG 123

QY 127 AGATCGAGGTGATCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 186
Db 124 AAGATCGAGGTGATCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 183

QY 187 TGATGTACGGCTCAAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGGACTACTTCAAG 246
Db 184 TGATGTACGGCTCAAGGCGCTTCATCAAGTACGTGTCCGGCATCCCGGACTACTTCAAG 243

QY 247 CAGTCCCTCCCGAGGGGCTTCACTGGAGCGGCAACCACTACGAGGACCGGGGCTTC 306
Db 244 CAGTCCCTCCCGAGGGGCTTCACTGGAGCGGCAACCACTACGAGGACCGGGGCTTC 303

QY 307 CTGACGGCCCAACAGGACACCTTCCCTGAGCGGCACTGCTGTGTACAAGTCAAGATC 366
Db 304 CTGACGGCCCAACAGGACACCTTCCCTGAGCGGCACTGCTGTGTACAAGTCAAGATC 363

QY 367 CTGGGCAACAACTTCCCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
Db 364 CTGGGTAACTTCCCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 423

QY 427 CCCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGCTGGCGGCGGCGGCGGCGGCGG 486
Db 424 CCGGAACCGAGATCGTGTACGAGGTGAGCGGCTGCTGGCGGCGGCGGCGGCGGCGG 483

QY 487 CTGAGGTGCCCCGGCGGCTGCGCACTGACCTGCGCACTGCAACCACTACCGTCCAAG 546
Db 484 CTGAAGTGCCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 543

QY 547 AAGCGCGCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 606
Db 544 AAGCGCGCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 603
```


Db 364 CTGTGTAATAATTTTCTGCTGATGGCCCGCTGATGAGAACAAAGACGAGAGATGGAG 423
 QY 427 CCTTCCACCGAGATCGTGTACAGAGTGACCGCGTGTGCGCGCCAGTCCCTGATGGCC 486
 Db 424 CCATCCACCGAGATAGTTTATGAAGTTGACGGTGTCTGCGTGGACAGTCTTTGATGGCC 483
 QY 487 CTGGAGTCCCGCGGCGTCCACCTGACCTGCGCACCTGCACACACCTACCGCTCCAAG 546
 Db 484 CTTAAGTCCCGTGGTGGTGTCTGATCTGACCTTGCATCTCCATCTACTTACAGTCCAAA 543
 QY 547 AAGCCCGCTCGCCCTGAAGATGCCCGCTTCCATCTCGAGGACCAACCGCATCGAGATC 606
 Db 544 AAACAGCTGTGCTTGAAGATGCCAGATTTTCAATTTGAAGATCATGCGATCGAGATA 603
 QY 607 CTGGAGGAGTGGAGAGGCGAAGTGTACAGAGTACAGGCGCGCGTGGGCGGTAC 666
 Db 604 ATGGAGGAGTGTGAGAAAGGCAAGTGTCTATAACAGTACGAAGCAGCAGTGGGCGGTAC 663
 QY 667 TCGCAGCGCGCCCTCCCAAGCTGGGCCACAAC 699
 Db 664 TGTGATGCTGTCCATCCAGCTTGGACATAC 696
 RESULT 5
 AX824727
 LOCUS AX824727 696 bp DNA linear PAT 11-DEC-2003
 DEFINITION Sequence 9 from Patent WO02068459.
 ACCESSION AX824727
 VERSION AX824727.1 GI:39750592
 KEYWORDS
 SOURCE Anemonia sulcata (snake-locks sea anemone)
 ORGANISM Anemonia sulcata
 Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
 Nynantheae; Actiniidae; Anemonia.
 REFERENCE 1
 AUTHORS Non aggregating fluorescent proteins and methods for using the same
 JOURNAL Patent: WO 02068459-A 9 06-SEP-2002;
 FEATURES Location/Qualifiers
 source 1..696
 /organism="Anemonia sulcata"
 /mol_type="unassigned DNA"
 /db_xref="taxon:6108"
 ORIGIN
 Query Match 62.3%; Score 440.2; DB 6; Length 696;
 Best Local Similarity 77.2%; Pred. No. 9e-46;
 Matches 535; Conservative 0; Mismatches 158; Indels 0; Gaps 0;
 QY 7 GCCTCCCTGTGACCGAGACCATGCCCTTACAGACCACTATGAGGCGACCGCTGAACGGC 66
 Db 4 GCCTCCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTAAGAGGCGAGTAAATGGC 63
 QY 67 CACTACTTCAAGTGCACCGGAGGCGGAGGCAACCCCTCGAGGCGACCCAGGAGATG 126
 Db 64 CACTACTTCAAGTGTACAGAAAGAGAGAGGCAACCCATTTGAGGGTACGAGGAATG 123
 QY 127 AAGATCAGGTGATCGAGGCGGCGCCCTGCTCCCTTCGCTTCCATCTGCTCCACCTCC 186
 Db 124 AAGATAGAGTTCATCGAAGGAGGTCCATTTGCCATTTTGCCTTCCACATTTTGTCAACGAGT 183
 QY 187 TGCATGACCGCTCCAGGCGCTTCACTCACTAGTGTGCGGCATCCCGACTACTTCAAG 246
 Db 184 TGTATGACGGTAGTAGAGGCTTCACTCACTAGTGTGCGGCATTTCTGACTACTTCAAG 243
 QY 247 CAGTCCCTCCCGAGGCTTCACTCGGAGCGCACCACTACGAGGACGCGCGCTTC 306
 Db 244 CAGTCTTCCCTGAAGGTTTTACTTGGGAAGAAGAACACACACCTACGAGGATGAGGCTTT 303
 QY 307 CTGACCGCCACACGAGACCTCCCTGGAGCGGCTGCTGCTGTACAGGTGAGATC 366
 Db 304 CTTACAGCTCATCAGGACACAGCCTAGATGGAGATTCCTCTGTTTACAGGTCAGATTT 363

QY 367 CTGGGCAACAACCTTCCCGCCGACGCGCCCGTGTGTCAGAACAAAGCGCGCGCTGGGAG 426
 Db 364 CTGTGTAATAATTTTCTGCTGATGGCCCGCTGATGAGAACAAAGACGAGAGATGGGAG 423
 QY 427 CCTTCCACCGAGATCGTGTACAGAGTGGACGCGTGTGCGCGCCAGTCCCTGATGGCC 486
 Db 424 CCATCCACCGAGATAGTTTATGAAGTTGACGGTGTCTGCGTGGACAGTCTTTGATGGCC 483
 QY 487 CTGGAGTCCCGCGGCGTGCACCTGACCTGCCACCTGCACACACCTACCGTCCCAAG 546
 Db 484 CTTAAGTCCCGTGGTGGTGTCTGATCTGACCTTGCATCTCCATCTACTTACAGTCCAAA 543
 QY 547 AAGCCCGCTCGCCCTGAAGATGCCCGCTTCCATCTCGAGGACCAACCGCATCGAGATC 606
 Db 544 AAACAGCTGTGCTTGAAGATGCCAGATTTTCAATTTGAAGATCATGCGATCGAGATA 603
 QY 607 CTGGAGGAGTGGAGAGGCGAAGTGTCTACAGAGTACAGGCGCGCGTGGGCGGTAC 666
 Db 604 ATGGAGGAGTGTGAGAAAGGCAAGTGTCTATAACAGTACGAAGCAGCAGTGGGCGGTAC 663
 QY 667 TCGCAGCGCGCCCTCCCAAGCTGGGCCACAAC 699
 Db 664 TGTGATGCTGTCCATCCAGCTTGGACATAC 696
 RESULT 6
 AF246709
 LOCUS AF246709 955 bp mRNA linear INV 29-AUG-2000
 DEFINITION Anemonia sulcata GFP-like chromoprotein FP595 mRNA, complete cds.
 ACCESSION AF246709
 VERSION AF246709.1 GI:9937257
 KEYWORDS
 SOURCE Anemonia sulcata (snake-locks sea anemone)
 ORGANISM Anemonia sulcata
 Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
 Nynantheae; Actiniidae; Anemonia.
 REFERENCE 1 (bases 1 to 955)
 AUTHORS Lukyanov, A.F., Fradkov, A.F., Gurskaya, N.G., Matz, M.V., Labas, Y.A., Savitsky, A.P., Markelov, M.L., Zaraisky, A.G., Zhao, X., Fang, Y., Tan, W. and Lukyanov, S.A.
 TITLE Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog
 JOURNAL J. Biol. Chem. 275 (34), 25879-25882 (2000)
 MEDLINE 20408921
 PUBMED 10852900
 REFERENCE 2 (bases 1 to 955)
 AUTHORS Matz, M.V., Fradkov, A.F., Lukyanov, K.A., Gurskaya, N.G. and Lukyanov, S.A.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAR-2000) Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya 16/10, Moscow 117871, Russia
 FEATURES Location/Qualifiers
 source 1..955
 /organism="Anemonia sulcata"
 /mol_type="mRNA"
 /db_xref="taxon:6108"
 CDS 69..767
 /note="purple chromoprotein; similar to green fluorescent protein"
 /codon_start=1
 /product="GFP-like chromoprotein FP595"
 /protein_id="AAG02385.1"
 /db_xref="GI:9937258"
 /translation="MASFLKKTMFPKFTTIGTNGHYFKCTKGEGNPFPGTQRMKIE VIEGGLPFAFHILSTSCMTGSKTFIKYSGIDYFKQSPFEGFTWERTTITDGGFL TAHQDTSLDGDCLVYKVLGNFPADGPNQKAGRWEPATEIIVEVDVLGQSILM ALKCPGGRHLTCHLHTTYSRKKPASALKMPGFHEDHIEIMBEVEKCKYQYEAAY GRYCDAAPSKLHN"
 ORIGIN
 Query Match 61.8%; Score 436.8; DB 3; Length 955;
 Best Local Similarity 76.7%; Pred. No. 2.2e-45;
 Matches 534; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 7 GCCTCCCTGCTGACGAGACATCCCTTCAGAGACCACTATCGAGGACCCGTTGAACGGC 66
 Db |||||
 72 GCTTCCTTTTAAAGAGACTATGCCCTTTAAAGACGACCAATGAAGGACGGTTAATGGC 131
 QY 67 CACTACTTCAAGTACACCGGCAAGGCGAGGCAACCCCTCGAGGGCACCCAGGAGATG 126
 Db |||||
 132 CACTACTTCAAGTACAGGAAGAGAGAGGCAACCCATTTGAGGGTACGAGGAATG 191
 QY 127 AAGATCAGAGTATCGAGGCGGCGCCCTGCTGCTTCACATCTGTCCTCCACTCC 186
 Db |||||
 192 AAGATAGAGGTCATCGAAGAGGTCCTATGCCATTTGCCCTTCACATTTGTCAACGAT 251
 QY 187 TGCATGACGGCTCCAAAGGCTTCATCAAGTACGTGTCGGGCTCCCGACTACTTCAAG 246
 Db |||||
 252 TGTATGACGCTAGTAAGACCTTCATCAAGTATGTGTACAGGATTCCTGACTACTTCAAG 311
 QY 247 CAGTCCCTCCCGAGGCTTCACCTGGAGGCGACACACCTTACAGAGCGGCGCTTC 306
 Db |||||
 312 CAGTCTTTCCTGAGGCTTTTACTTGGGAAGAACCAACCTTACGAGGATGAGGCTTT 371
 QY 307 CTGACCGCCACCAAGGACACCTCCCTGGACGGGACTGCTGCTGTACAGGTTGAAGATC 366
 Db |||||
 372 CTTACAGTCTATCAGGACACAAAGCTAGATGGAGATTGCTCTTTTACAGGTCAAGATT 431
 QY 367 CTGGGCAACACTTCCCGCGGAGCGCCCTGATGACAGAAAGCGCGCGCTGGAG 426
 Db |||||
 432 CTTGGTAATATTTCTGCTGATGCGCCCGTGTGACAGAACAAAGCAGGAAGTGGGAG 491
 QY 427 CCCTCCACCGAGATCGTGTACGAGTGGACGGGTGCTGCGCGGCGAGTCCCTGATGGC 486
 Db |||||
 492 CCAGCCACCGAGATAGTTATGAAGTGAAGGCTGCTGCTGGTGAAGCTTTTGTGGCC 551
 QY 487 CTGAGTGCCTCCCGGCTGCGACCTGACCTGCACTGCAACACCTACCGCTCCAG 546
 Db |||||
 552 CTTAAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 611
 QY 547 AAGCCCGCTCCCGCTGAGAGTCCCGGCTTCCACTTCGAGGACCAACGCTGAGATC 606
 Db |||||
 612 AAACAGCTAGTGCCTTGAAGTGCAGGATTTTCATTTGAAGATCATCGATCGAGATA 671
 QY 607 CTGAGAGGTGAGGAAGGCAAGTGTCAACAGTACGAGCGCGCTGGCGGCTAC 666
 Db |||||
 672 ATGAGGAAGTGAAGGAAGCAAGTGTCTATAACAGTACGAGCAGTGGGCGAGGTAC 731
 QY 667 TGCACCGCGCGCCCTCCAAAGCTGGGCAACACTGA 702
 Db |||||
 732 TGTGATGCTGCTCATCCAGCTTGCACATACTAA 767

RESULT 7
 AF322222
 LOCUS
 DEFINITION Anemonia sulcata nonfluorescent red protein asCP562 mRNA, complete cds.
 ACCESSION AF322222
 VERSION AF322222.2 GI:18999345
 KEYWORDS
 SOURCE
 ORGANISM
 Anemonia sulcata (snake-locks sea anemone)
 Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
 Nynanthaeae; Actiniidae; Anemonia.
 1 (bases 1 to 699)
 Wiedemann, J., Elke, C., Spindler, K.D. and Funke, W.
 Cracks in the beta-can: fluorescent proteins from Anemonia sulcata
 (Anthozoa, Actiniaria)
 Proc. Natl. Acad. Sci. U.S.A. 97 (26), 14091-14096 (2000)
 20570437
 MEDLINE
 PUBLISHED 1121018
 2 (bases 1 to 699)
 Wiedemann, J., Elke, C., Spindler, K.-D. and Funke, W.
 Direct Submission
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Universitaet Ulm, Albert-Einstein-Allee 11, Ulm 89069, Germany
 3 (bases 1 to 699)
 Wiedemann, J., Elke, C., Spindler, K.-D., Funke, W., Sundin, B. and Jach, G.
 Direct Submission
 Submitted (28-FEB-2002) Allgemeine Zoologie und Endokrinologie,
 Universitaet Ulm, Albert-Einstein-Allee 11, Ulm 89069, Germany
 Sequence update by submitter
 On Feb 28, 2002 this sequence version replaced gi:11890752.
 Location/Qualifiers
 1..699
 /organism="Anemonia sulcata"
 /mol_type="mRNA"
 /db_xref="taxon:6108"
 1..699
 /notes="similar to GFP"
 /product="nonfluorescent red protein asCP562"
 /protein_id="AAG1206.2"
 /db_xref="GI:18999346"
 /translation="MASFLKTKMPFKTTIEGTVNGHYFKCTGKGNPEFTQEMKIE
 VIRGGPLPFAFHILSTCMVGSKTPIKVVSGIPDYFKQSPFSGFTWERTTTTVDGGL
 TADQTSLDGDCIVKVKILGNFPADGPMONKAEKRPATETLLYEVDGVLRGSLM
 ALKCPGGRHLTCHLHSTYRSKRPASALNKPFGPHDRHRIEINEEVEKCKYQYAAV
 ARYCDAPSKLGH"

ORIGIN

Query Match 60.7%; Score 428.8; DB 3; Length 699;
 Best Local Similarity 76.0%; Pred. No. 2.4e-44;
 Matches 529; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
 QY 7 GCCTCCCTGCTGACGAGACCACTCCCTTCAGGACCACTATCGAGGACCCGTTGAACGGC 66
 Db |||||
 4 GCTTCCTTTTAAAGAGACTATGCCCTTTAAAGACGACCAATGAAGGACGGTTAATGGC 63
 QY 67 CACTACTTCAAGTACACCGGCAAGGCGAGGCAACCCCTCGAGGGCACCCAGGAGATG 126
 Db |||||
 64 CACTACTTCAAGTACAGGAAGAGAGAGGCAACCCATTTGAGGGTACGAGGAATG 123
 QY 127 AAGATCAGAGTATCGAGGCGGCGCCCTGCTGCTTCACATCTGTCCTCCACTCC 186
 Db |||||
 124 AAGATAGAGTATCGAAGGAGGTCCATTTGCCCTTCACATTTGTCAACGAGT 183
 QY 187 TGCATGACGGCTCCAAAGGCTTCATCAAGTACGTGTCGGGCTCCCGACTACTTCAAG 246
 Db |||||
 184 TGTATGACGCTAGTAAGACCTTCATCAAGTACGTGTCAGGATTCAGACTACTTCAAG 243
 QY 247 CAGTCCCTCCCGAGGCTTCACCTGGAGCGGACCAACCTTACGAGACGCGCGCTTC 306
 Db |||||
 244 CAGTCTTTCCTGAGGCTTTTACTTGGGAAGAACCAACCTTACGAGGATGAGGCTTT 303
 QY 307 CTGACCGCGCCACCAAGGACCACTCCCTCGAGCGGAGTGCCTGCTGTACAGGTGAAGATC 366
 Db |||||
 304 CTTACAGCTCATCAGGACACAGCCCTAGATGGAGATTGCTCTGTTTACAGGTCAAGATT 363
 QY 367 CTGGGCAACAACTTCCCGCGGAGCGCCCTGCTGATGTCAGAAACAGCGCGCGCTGGAG 426
 Db |||||
 364 CTTGGTAATATTTTCTGCTGATGCGCCCGTGTGACAGGACCAACAAAGCAGGAAGTGGAG 423
 QY 427 CCCTCCACCGAGATCGTGTACGAGGTCGAGCGCTGCTGCGGGCGCAGTCCCTGATGGC 486
 Db |||||
 424 CCAGCCACCGAGATCTTATGAAGTGAAGGCTGCTGCTGGTGAAGCTTTTGTATGGCC 483
 QY 487 CTGAGTGCCTCCCGGCTGCGACCTGACCTGCCACCTGTGACACCACTTACCGCTCCAG 546
 Db |||||
 484 CTTAAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
 QY 547 AAGCCCGCTCCCGCTGAGAGTGCCTGCTTCCACTTCGAGGACCAACGCTGAGATC 606
 Db |||||
 544 AAACAGCTAGTGCCTTGAAGTGAAGGATTTTCATTTTGGAGATCATCGCATCAGATA 603
 QY 607 CTGAGAGGTGAGGAAGGCAAGTGTCTATAACAGTACGAGCGCGCTGGCGGCTAC 666
 Db |||||

QY 552 CGCTCCGCTGAAGATGCGCGCTTCCACTTCGAGGACACCGCATCGAGATCCTGGA 611
 |||
 Db 549 CGTGGGGCCCTGACCATGCCGGCTTCCACTTCACCGACATCCCGCTGCAGATCCTGG 608
 |||
 QY 612 GGAGGTGGAGAAAGGCAAGTGTACAAGCAGTACAGGCGCGCTGCGGTACTTGCGA 671
 |||
 Db 609 GAA---GGAGAAGGACGAGTACTTCGAGCTGTACGAGGCCAGGTGGCCCGGTACAGGA 665
 |||
 QY 672 C 672
 |||
 Db 666 C 666

RESULT 10
 AX527902 AX527902 1396 bp DNA linear PAT 21-NOV-2002
 LOCUS
 DEFINITION Sequence 15 from Patent WO0230965.
 ACCESSION AX527902
 VERSION AX527902.1 GI:25172350
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.

REFERENCE
 1
 LUKYANOV, S.A., FRADKOV, A.F., LUKYANOV, K.A. and GURSKAYA, N.G.
 Nucleic acids encoding stichodactylidae chromoproteins
 Patent: WO 0230965-A 15 18-APR-2002;
 Clontech Laboratories Inc. (US)

FEATURES
 source
 1..1396
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="fusion construct"

ORIGIN

Query Match 59.1%; Score 417.8; DB 6; Length 1396;
 Best Local Similarity 79.1%; Pred. No. 4.5e-43;
 Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;

QY 12 CTTGCTGACCGAGACCATGCTTTCAGGACCACTCGAGGCGACCGTGAACGGCCACTA 71
 |||
 Db 25 CTTGCTGAAGAGAGATGCGCATCAAGATGTATGAGGCGACCGTGAACGGCCACTA 84
 |||
 QY 72 CTTCAAGTGCACCGGCAAGGGGAGGCGAACCCCTTCGAGGCGACCCAGGAGATGAAGT 131
 |||
 Db 85 CTTCAAGTGCAGGGGAGGCGAGCGCAACCCCTTCGCGCGCACCCAGAGCATCGGAT 144
 |||
 QY 132 CGAGGTGATCGAGGGGGGGCCCTTCGCTTCGCTTCCACATCTCTGCTCCACCTCTGAT 191
 |||
 Db 145 CCACGTGACGAGGGGGGGCCCTTCGCTTTCGATCTCTGCGCTTCGATCTCTGCTGCGA 204
 |||
 QY 192 GTACGGCTCCAAGGCTTCATCAAGTACGTGTCGGGCATCCCGCATCTTCAAGCAGTC 251
 |||
 Db 205 GTAGGCGACGAGACCTTCGTGCACACACCGCGAGATCCCGCATCTTCAAGCAGAG 264
 |||
 QY 252 CTTCCCGAGGGTTTCACTGGAGCGGACACACCTACGAGGAGCGGGCTTCTGATC 311
 |||
 Db 265 CTTCCCGAGGGTTTCACTGGAGAGAACCCACCTACGAGGAGCGGGCATCTGATC 324
 |||
 QY 312 CGCCACACGAGCACCTTCCTGAGCGGCGACTGCTGTGTACAAGTGAAGTCTGCG 371
 |||
 Db 325 CGCCACACGAGCACCTTCCTGAGGGGAACTGCTGTATCTACAAGTGAAGTCTGCG 384
 |||
 QY 372 CAACAACTTCCCGCGCGCGCCCGTGTATGAGAGAACAGGCGCGCGCTGGAGCCCTC 431
 |||
 Db 385 CACCAACTTCCCGCGCGCGCGCGCTGTATGAAGAACAGAGCGCGGGCTGGAGCCAG 444
 |||
 QY 432 CACCGAGATCTGTACAGGTGAGCGGCTGTGCGCGGCGAGTCCCTGATGCGCCTGGA 491
 |||
 Db 445 CACCGAGGTGGTGTACCCGAGAACCGGCTGTGTGTCGCGCGGAGACGTATGGCCCTGAA 504
 |||
 QY 492 GTGCCCCGGGCTGCGCACTGACCTGCCACCTGACACACCTACCGCTCCAAGAGGCC 551
 |||

Db 505 G---GTGGGCAACCGCGGCTGATCTCCACCATACACAGTACCGGAGCAAGAGGC 561
 |||
 QY 552 CGCTCCCGCTGAAGATGCCCGCTTCCACTTCGAGGACCAACCGCATCGAGATCCTGGA 611
 |||
 Db 562 CGTGGGGCTGACCATGCCCGCTTCCACTTCACGACATCCGGCTGCAGATGCTGCG 621
 |||
 QY 612 GGAGGTGGAAGGCAAGTGTACTACAGCAGTACGAGGCGCGCTGGGCGCTACTCTCGA 671
 |||
 Db 622 GAA---GGAGAAGGACGAGTACTTCGAGCTGTACGAGCGACGCTGCGCCGCTACAGGA 678
 |||
 QY 672 C 672
 |||
 Db 679 C 679

RESULT 11
 AX527904 AX527904 1424 bp DNA linear PAT 21-NOV-2002
 LOCUS
 DEFINITION Sequence 17 from Patent WO0230965.
 ACCESSION AX527904
 VERSION AX527904.1 GI:25172351
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.

REFERENCE
 1
 LUKYANOV, S.A., FRADKOV, A.F., LUKYANOV, K.A. and GURSKAYA, N.G.
 Nucleic acids encoding stichodactylidae chromoproteins
 Patent: WO 0230965-A 17 18-APR-2002;
 Clontech Laboratories Inc. (US)

FEATURES
 source
 1..1424
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="fusion construct"

ORIGIN

Query Match 59.1%; Score 417.8; DB 6; Length 1424;
 Best Local Similarity 79.1%; Pred. No. 4.5e-43;
 Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;

QY 12 CTTGCTGACCGAGACCATGCTTTCAGGACCACTCGAGGCGACCGTGAACGGCCACTA 71
 |||
 Db 25 CTTGCTGAAGAGAGATGCGCATCAAGATGTATGAGGCGACCGTGAACGGCCACTA 84
 |||
 QY 72 CTTCAAGTGCACCGGCAAGGGGAGGCGAACCCCTTCGAGGCGACCCAGGAGATGAAGT 131
 |||
 Db 85 CTTCAAGTGCAGGGGAGGCGAGCGCAACCCCTTCGCGCGCACCCAGAGCATCGGAT 144
 |||
 QY 132 CGAGGTGATCGAGGGGGGGCCCTTCGCTTTCGCTTCCACATCTCTGCTCCACCTCTGAT 191
 |||
 Db 145 CCACGTGACGAGGGGGGGCCCTTCGCTTTCGATCTCTGCGCTTCGATCTCTGCTGCGA 204
 |||
 QY 192 GTACGGCTCCAAGGCTTCATCAAGTACGTGTCGGGCATCCCGCATCTTCAAGCAGTC 251
 |||
 Db 205 GTACGGGAGGAGGCTTCGTCGACACCGCGAGATCCCGCATCTTCAAGCAGAG 264
 |||
 QY 252 CTTCCCGAGGGTTTCACTGGGAGCGCACACCTACGAGGAGCGGGCTTCTCTGAC 311
 |||
 Db 265 CTTCCCGAGGGTTTCACTGGGAGAGAACCCACCTACGAGGAGCGGGCATCTCTGAC 324
 |||
 QY 312 CGCCACACGAGCACCTTCCTGAGCGGCGACTGCTGTGTACAAGTGAAGTCTGCG 371
 |||
 Db 325 CGCCACACGAGCACCTTCCTGAGGGGAACTGCTGTATCTACAAGTGAAGTCTGCG 384
 |||
 QY 372 CAACAACTTCCCGCGCGCGCCCGTGTATGAGAGAACAGGCGCGCGCTGGAGCCCTC 431
 |||
 Db 385 CACCAACTTCCCGCGCGCGCGCGCTGTATGAAGAACAGAGCGCGCGCTGGAGCCAG 444
 |||
 QY 432 CACCGAGATCTGTACAGGTGAGCGGCTGTGCGCGGCGAGTCCCTGATGCGCCTGGA 491
 |||


```

Db      658  TTATCGGGGTGAACATTCCTCCGACGCGCCCGTGTATCAGAGAAGACCATGGGCTGG 717
Qy      424  GAGCCCTCCACCGAGATCGTACGAGGTGGACGGGTGCTGCGGGCCAGTCCCTGATG 483
Db      718  GAGGCTTCACCGAGGCGCTGTACCCCGGACGGGTGCTGAAGGGGAGACCCACAAG 777
Qy      484  GCGCTGAGTGCGCCCGGGGTGCGCACCTGACCTGCCACCTGACACCACTACCGTCC 543
Db      778  GCGCTGAAGTGAAGACGCGCGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCC 837
Qy      544  AAGAAGCCCGCTCCCGCTGAGATGCGCGGTTCCACTTCGAGGACCAACCGCATCGAG 603
Db      838  AAGAAGCCCG-----TGCAGCTGCCCGGCTACTACTAGTGACCGCAAGCTGAC 888
Qy      604  ATCCTGGAGAGTGGAGAGGCAAGTGTCTAAGCAGTACGAGCGCCCGTGGCGCCG 663
Db      889  ATCACTCCCAACAGAGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCG 948
Qy      664  TAC 666
Db      949  CAC 951

RESULT 14
AX370408
LOCUS      AX370408      678 bp      DNA      linear      PAT 16-FEB-2002
DEFINITION      Sequence 5 from Patent WO0196373.
ACCESSION      AX370408
VERSION      AX370408.1 GI:18857492
KEYWORDS      .
SOURCE      synthetic construct
ORGANISM      synthetic construct
              artificial sequences.
REFERENCE      1
AUTHORS      Pradkov,A.F. and Tersikh,A.
TITLE      Fluorescent timer proteins and methods for their use
JOURNAL      Patent: WO 0196373-A 5 20-DEC-2001;
              Clontech Laboratories Inc. (US)
FEATURES      Location/Qualifiers
              source
              1..678
              /organism="synthetic construct"
              /mol_type="unassigned DNA"
              /db_xref="taxon:32630"
              /note="variant of sequence from Discosoma sp."

ORIGIN
Query Match      44.3%; Score 313.4; DB 6; Length 678;
Best Local Similarity 68.3%; Pred. No. 5.1e-30;
Matches 453; Conservative 0; Mismatches 201; Indels 9; Gaps 1;

Qy      4  TCCGCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAAC 63
Db      10  TCCGAGAACGTATCATCCGAGTTTCATGCGTTCAAGGTGCGCATGGAGGCGACCGTGAAC 69
Qy      64  GGCCTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTTCGAGGCGACCCAGGAG 123
Db      70  GGCACAGATTGAGATCGAGGCGAGGGCGAGGGCGCGCCCTTACGAGGCGCCACAAC 129
Qy      124  ATGAGATCGAGTGATCGAGGCGCGCCCTGCTGCTTCCATCTCCATCTCCATCTCC 183
Db      130  GTGAAGTGAAGTGAACAAGGGCGCGCCCTGCTGCTTCCGCTGGGACATCTCTTCC 189
Qy      184  TCCTGCTATGACGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTC 243
Db      190  CAGTTCAGTACGCTCCAAAGTGTACGTGAAGCACCCCGCGACATCCCGCACTACAAG 249
Qy      244  AAGCAGTCCCTCCCGAGGGCTTCACTGGAGGCGACCACTACGAGGAGCGGGCG 303
Db      250  AAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGTGAATTCGAGGACGGCGCG 309
Qy      304  TTCCTGACCGCCACAGGACACTCTCCTTGAACGCGCACTGCTGTGTATCAAGGTGAAG 363

```

```

Db      310  GTGCGGACCGTGACCCAGGACTCTCTCCCTGAGGACGGCTTCTATCTACAAGGTGAAG 369
Qy      364  ATCTGGGCAACAATTTCCCGCGGACGCGCCCGTGTATGAGAACAGAGCGCGCGCTGG 423
Db      370  TTCATCGCGGTGAATTTCCCTCCGACGCGCCCGTGTATGAGAGAGAACCATGGGCTGG 429
Qy      424  GAGCCCTCCACCGAGATCGGTAGAGGTGAGCGGCTGCTGCGCGCCAGTCCCTGATG 483
Db      430  GAGGCTTCACCGAGCGCTGTACCCCGGCAAGCGCTGCTGAAGGGGAGATCCACAAG 489
Qy      484  GCGCTGAGTSCCCCGCGGTGCGCACCTGACCTGCCACCTGCGACACCACTACCGTCC 543
Db      490  GCGCTGAAGTGAAGGACGCGCGCCACTACTCTGGTGGAGTTCAAGTCCATCTACATGGCC 549
Qy      544  AAGAAGCCCGCTCCCGCTGAGATGCGCGGTTCCACTTCGAGGACCAACCGCATCGAG 603
Db      550  AAGAAGCCCG-----TGCAGCTGCCCGGCTACTACTAGTGAGCACCAACTGGAC 600
Qy      604  ATCTGGAGAGTGGAGAGGCAAGTGTCTAAGCAGTACGAGGCGCGCGTGGCGCGC 663
Db      601  ATCACTCCCAACAGGAGACTACACCATCTGTGAGCAGTACGAGCGCACCGAGGCGCG 660
Qy      664  TAC 666
Db      661  CAC 663

RESULT 15
AX824732
LOCUS      AX824732      678 bp      DNA      linear      PAT 11-DEC-2003
DEFINITION      Sequence 14 from Patent WO02068459.
ACCESSION      AX824732
VERSION      AX824732.1 GI:39750595
KEYWORDS      .
SOURCE      synthetic construct
ORGANISM      synthetic construct
              artificial sequences.
REFERENCE      1
AUTHORS      Non aggregating fluorescent proteins and methods for using the same
TITLE      Patent: WO 02068459-A 14 06-SEP-2002;
JOURNAL      Location/Qualifiers
FEATURES      source
              1..678
              /organism="synthetic construct"
              /mol_type="unassigned DNA"
              /db_xref="taxon:32630"
              /note="non-aggregating mutant"

ORIGIN
Query Match      44.3%; Score 313.4; DB 6; Length 678;
Best Local Similarity 68.3%; Pred. No. 5.1e-30;
Matches 453; Conservative 0; Mismatches 201; Indels 9; Gaps 1;

Qy      4  TCCGCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAAC 63
Db      10  TCCGAGAACGTATCATCCGAGTTTCATGCGTTCAAGGTGCGCATGGAGGCGACCGTGAAC 69
Qy      64  GGCCTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTTCGAGGCGACCCAGGAG 123
Db      70  GGCACAGATTGAGATCGAGGCGAGGGCGAGGGCGCGCCCTTACGAGGCGCCACAAC 129
Qy      124  ATGAGATCGAGTGATCGAGGCGCGCCCTGCTGCTTCCATCTCCATCTCCATCTCC 183
Db      130  GTGAAGTGAAGTGAACAAGGGCGCGCCCTGCTGCTTCCGCTGGGACATCTCTTCC 189
Qy      184  TCCTGCTATGACGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTC 243
Db      190  CAGTTCAGTACGCTCCAAAGTGTACGTGAAGCACCCCGCGACATCCCGCACTACAAG 249
Qy      244  AAGCAGTCCCTCCCGAGGGCTTCACTGGAGGCGACCACTACGAGGAGCGGGCG 303
Db      250  AAGCTGTCTTCCCGAGGGCTTCAAGTGGAGCGCGTGTGAATTCGAGGACGGCGCG 309

```

QY 304 TTCCTGAGCGCCACCAGGACACCTCCCTGGACGGCGACTGCTGGTGTACAAGGTGAAG 363
Db |||||
QY 310 GTGGGACCGGTGACCCAGGACTCTCCCTGCTGAGGACGGCTGCTTCATCTACAAGGTGAAG 369
Db |||||
QY 364 ATCCTGGGACACAACTTCCCGCCGAGACGGCCCGGTGATGACAGAAACAAGGCCGGCCGCTGG 423
Db |||||
QY 370 TTCATCGGCGTGAATTCCTCCCTCCGACGGCCCGGTGATGACAGAAACAAGGCCGGCTGG 429
Db |||||
QY 424 GAGCCCTCCACCGAGATCGTGTACGAGGTGGACGGCGTGTGCGGCGCCAGTCCCTGATG 483
Db |||||
QY 430 GAGGCTTCACCGAGCGCTGTACCCCGGACGGCGTGTGAGGGGAGATCCACAAG 489
Db |||||
QY 484 GCCCTGGAGTCCCGCGGTCGCCACCTGACCTGCCACCTGCCACACCACTACCGTCC 543
Db |||||
QY 490 GCCCTGAAGCTGAAGGACGGCGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCC 549
Db |||||
QY 544 AAGAGCCCGCTCCGCCCTGAAGATGCCGGCTTCCACTTCGAGGACCAACCGCATCGAG 603
Db |||||
QY 550 AAGAGCCCG-----TGCAGTCCCGGCTACTACTACGTGGACACCAAGCTGGAC 600
Db |||||
QY 604 ATCCTGGAGGAGTGGAGAGGGAAGTGTCTACAAGCAGTACGAGGCCCGCTGGGCCGC 663
Db |||||
QY 601 ATCACCTCCACACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGGC 663
Db |||||
QY 664 TAC 666
Db |||||
QY 661 CAC 663

Search completed: August 3, 2004, 14:04:19
Job time : 4732.5 secs

Db 327 GGAGATCCGACGACATCAACCTGATCGAGCAAGTTCTGTACCGGTGGAGTACAA 386
Qy 369 GGGCAACAATTCCTCCGCGACGCGCCCGTGTATGATGAGAACAAAGCCGCGCTGGAGCC 428
Db 387 GGGCAGCAATTCCTCCGCGACGCGCGCGTGTATGATGAGAACCAATTCCTGGGATCGAGCC 446
Qy 429 CTCACCCAGATCGTGTACGAGGTGGAGCGGCGTGTCTGCGCGGCCAGTCCCTGATGGCCCT 488
Db 447 CAGCTTCAGGCCATGTATCATGAACAAAGCGGTCTGTGGGCGAGGTATCTCTGTGTGTA 506
Qy 489 GGAGTCCCCCGCGTCTGCCACCTGACCTGCCACCTGACACACACCTACCGCTCCAAAGAA 548
Db 507 CAAGCTGAACAGCGCAGTACTACAGCTGCCACATGAGAC-----CTGTGAGAGAG 560
Qy 549 GCGCCCTCCGCGCTGAAGATGCGCGGCTTCCACTTTCGAGGACCAACCGCATCGAGATCCT 608
Db 561 CAAGGCGGTGTGAAGGAGTTCCCTCTCATCCACTTCATCCAGACCGCGCTGGAGA---A 617
Qy 609 GGAGGAGTGGAGAGGCAAGTGTACAGCAGTACGAGCGCGCGTGGGCC 661
Db 618 GACCTACGTGAGAGCGCGCTTCTGTGGAGCAGCAGAGCCGCCATCGCCC 670

RESULT 2

US-09-459-956-7
; Sequence 7, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1230-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Clavularia sp
US-09-459-956-7

Query Match 23.3%; Score 165; DB 4; Length 801;
Best Local Similarity 54.9%; Pred. No. 1.7e-24;
Matches 352; Conservative 0; Mismatches 280; Indels 9; Gaps 1;
Qy 27 CATGCCCTTCAGGACCACTCGAGGCGACCGTGAAGCGGCACCTCAAGTGCACCGG 86
Db 147 CATGAAGATTAACTGAAGTGAAGGAATGTAACCGGCATGCTTTGTGTATCGAAG 206
Qy 87 CAAGGGCGAGGGCAACCCCTCGAGGGCACCAGGAGATGAAGATCGAGGTGATCGAGG 146
Db 207 AGNAGGAAGAAAGCCCTTACGATGGGACACACACTTTAAACCTGGAAGTGAAGGAG 266
Qy 147 CGGCGCCCTGCGCTTCGCTCCACATCCTGTCCACCTCTGTCATGATACCGCTCCAGGC 206
Db 267 TCGCGCTCTGCTTTTCTTACGATATCTTGTCAAAAGCGGTTCAGTACCGAAACAGAGC 326
Qy 207 CTTCATCAAGTACGTGTCCGCGATCCCGACTTCTCAAGCAGTCCCTCCCGAGGGCTT 266
Db 327 ATTGACAAAATACCCAGAGATATAGCAGACTATTTCAAGCAGTCTGTTCCCGAGGATA 386
Qy 267 CACCTGGAGGCGACACCACTAGGAGGCGCGCTTCTGTACCCGCCACACAGGACAC 326
Db 387 TTCCTGGGAAGAACCACTGACTTTTGAAGAACAAAGGCAATTTGTAAGTGAAGAGTGCAT 446

Qy 327 CTCCTTGGACGCGACTGCTGCTGTACAGGTGAGATCTTGGACACACTTCCCGC 386
Db 447 AAGCATGGAGGAAGACTCTCTTTATCTATGAATTCGTTTGTATGGATGAACCTTTCCTCC 506
Qy 387 CGACGCGCCCGTGTATGATGAGAACAAAGCGCGCTGAGGAGCCCTCCACCGAGATCGTGA 446
Db 507 CAATGGTCCGTTATGATGAGAAAACCTTGAAGTGGGAACCATCCATCGATGATATGA 566
Qy 447 CGAGTGGACGCGCTGCTGCGGCGCAGTCCCTGTATGCGCCCTGAGATGCGCGCGCTG 506
Db 567 CGTGGCTGATGAGTCTGCTGCGAGATATTAGCCATTTCTCTGTTGCTGGAGGAGGTGG 626
Qy 507 CCACCTGACTCCACCTGACACACCTACCGCTCCAAAGAGCCGCTCCGCGCTCAA 566
Db 627 CAATTACGATGTGACTTCAAAAGTATTTACAAGCAAAAAA-----GTTGTCAA 677
Qy 567 GATGCCCGCTTCCACTTTCGAGGACCAACCGCATCGAGATCTTCGAGGAGGTGGAGAGG 626
Db 678 ATTGCCAGACTATCACTTTGTGACCATCGCATTCGATTCGATGATCTTGAACCATGACAGGATTA 737
Qy 627 CAAGTGTCTACAGCAGTACGAGGCGCGCTGGCGCTACT 667
Db 738 CAACAAAGTACGCTGTATGAGATGTCAGTTCTCTCGCTATT 778

RESULT 3

US-09-459-956-6
; Sequence 6, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1230-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-09-459-956-6

Query Match 22.5%; Score 159.4; DB 4; Length 678;
Best Local Similarity 56.1%; Pred. No. 2.2e-23;
Matches 301; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
Qy 17 TCACCGACACATGCTCCCTTCAGGACCACTCGAGGCGACCGTGAACGCGCACTACTTCA 76
Db 23 TCAAGAGTTTCATGAGTTTAAAGTTTCGATGGAAGAACGTCATGCGCAGGATTTG 82
Qy 77 AGTGACCGCGCAAGGCGGAGGCGCAACCCCTCGAGGGCACCAGGAGATGAAGATCGAGG 136
Db 83 AATAGAAGGCGAAGGAGAGGAGGAGGCGCATACGAAGGCCACATACCGTAAAGTTAAG 142
Qy 137 TGATGAGGCGGCGCCCTGCGCTTCGCTTCACATCTCTCCACCTCTCCATGTCATG 196
Db 143 TAACCAAGGCGGAGCGCTTTCGCAATTTGCTGGATATTTTGTCCACCAATTTTCAGTATG 202
Qy 197 GCTTCAAGGCGCTTCATCAAGTACGTGTCCGCGCATCCCGACTTCTCAAGCAGTCCCTCC 256
Db 203 GAAGCAAGGTATATGTCAGACACCTTCGCGACATACAGACTATAAAGCTGTCATTTC 262
Qy 257 CGAGGGCTTACCTCGGAGGCGCACCACTACGAGGACGGCGCTTCTTGACCGGCC 316

Db 263 CTGAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAAGACGGTGGCGTCGTACTGTAA 322
Qy 317 ACCAGGACACTCCCTGGACGGGACTGCTGTGTACAAAGGTGAAGATCCTGGGCAACA 376
Db 323 CCCAGGATCCAGTTTGCAGGATGCTGTTTCATCTACAAGGTCAAGTTCAITGGCGTGA 382
Qy 377 ACTTCCCGGCGAGCGCCCGTGTATCAGAAACAGGCCCGCGCTGGAGCCCTCCACCG 436
Db 383 ACTTTCCTTCCGATGACCTGTTATGCAAAAGAGACAATGGGCTGGGAAGCGACACTG 442
Qy 437 AGATCGTGTACGAGGTGGACGGGTGCTGCGCGCCAGTCCCTGTATGGCCCTGGAGTGCC 496
Db 443 AGGTTTGTATCTCTGTATGGGTGTTGAAAGAGAGATTTCATAAGGCTCTGAGCTGA 502
Qy 497 CCGCGGTGCCACCTGACCTGCCACCTGCACACCACTACCGCTCCAAAGACCGG 553
Db 503 AAGACGGTGTCTATTACCTAGTTGAATTCAAAAGTATTTACATGGCAAAAGACCTG 559

RESULT 4

US-09-459-956-5
; Sequence 5, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 699
; TYPE: DNA
; ORGANISM: *Drosophila striata*
US-09-459-956-5

Query Match 21.9%; Score 154.8; DB 4; Length 699;
Best Local Similarity 54.4%; Pred No. 1.8e-22;
Matches 337; Conservative 0; Mismatches 277; Indels 6; Gaps 1;
Qy 47 TCGAGGCGACCGTGAACGGCCACTTCAAGTGCACCGCAAGGGCGAGGCAACCCCC 106
Db 53 TGGAGGAAACGTTCAATGGGCACTTCTGAATATAAAGGCAAGGAAAGGACAGCCTA 112
Qy 107 TCGAGGCGACCGAGGATGAAGATCGAGGTATCGAGGGCGCCCTCGCCTCGCCT 166
Db 113 ATGAAGGCAACCAATACCGTCACGCTCAGGTGTACCAAGGTGACCTCTCCCAATTTGGTT 172
Qy 167 TCCACATCTCTGCCACCTCTGATGACGGCTCCAAGGCTTCATCAAGTACGTGTCCG 226
Db 173 GGCATATTTGTGCGCCCAATTTTCAGTATGGAACAGGCAATTTGTCACCACTGACA 232
Qy 227 GCATCCCGCACTATTCAAGCAGTTCCTCCCGAGGGCTTCACTGGGAGCGGCAACCA 286
Db 233 ACATACATGATTATCTAAAGCTGTCTATTCGGAGGGATATACATGGGAACGGTCCATGC 292
Qy 287 CCTACGAGGACGGCGCTTCTGTACCGCCCAACAGGACACTCCCTCGAGCGGCACTGCC 346
Db 293 ACTTTGAAGACGGTGGCTTGTGTGTATCACCATGATATCAGTTTGACAGGCACTGTT 352
Qy 347 TGGTGTACAAGGTGAAGATCTGGGCAACAACTTCCCGCGGAGCGGCCCTCGATGAGA 406
Db 353 TCTACTACGACATCAAGTTCACCTGGCTTGAATTTCTCCAAATGGACCGGCTGTGCGAGA 412

Qy 407 ACAAGGCGGCGCGTGGAGCCCTCCACGAGATCGTGTACGAGGTGAAGCGGTCTGTC 466
Db 413 AGAAGCAAACTGGCTGGGAACCGAGCACTGAGCCTTTGTATCTCGTGTATGTTGTA 472
Qy 467 CCGGCGAGTCCCTGATGGCCCTGGAGTGCCCGCGGTGCGCCACTGACCTGACCTGCC 526
Db 473 TAGGAGACATCATCATGCTCTGACAGTTGAAGAGGTGGTCAATACGATGTGACATTA 532
Qy 527 ACACCACTACGCTCCAGAAAGCCGCTCCGCGCTGAGATGCGCGGCTTCCACTTCG 586
Db 533 AAATCTGTTTACAGGCGCAAGAAAGCCGCT-----TGAAGATCCAGGGTATCACATG 586
Qy 587 AGGACCAACCGCATCGAGATCCTGAGGAGGTGGAGAAAGGCAAGTGTCTACAAGCAGTACG 646
Db 587 TTGACACAAACTGGTTATATGGAACAACGACAAAGATTCATGAAGTTGAGGAGCATG 646
Qy 647 AGGCGCGCGTGGCGCGCTAC 666
Db 647 AAATCGCGTTGCGACGCCAC 666

RESULT 5

US-09-459-956-2
; Sequence 2, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 690
; TYPE: DNA
; ORGANISM: *Anemonia majano*
US-09-459-956-2

Query Match 17.1%; Score 121; DB 4; Length 690;
Best Local Similarity 51.8%; Pred No. 9.3e-16;
Matches 333; Conservative 0; Mismatches 295; Indels 15; Gaps 2;
Qy 22 GAGACCATGCCCTTCAGGACCAACCATCGAGGGCACCGTGAACGGCCACTACTTCAAGTGC 81
Db 28 GATGACATGAAATGACCTACCATATGGATGGCTGTGTCAATGGGCATTACTTTACCGTC 87
Qy 82 ACCGCAAGGGGAGGGCAACCCCTCGAGGGCACCA-----GGAGATGAAGATCGAG 135
Db 88 AAAGGTGAAGGCAACGGGAAGCCATACGAAGGGAGCGAGACTTCGACTTTTAAAGTACC 147
Qy 136 GTGATCGAGGGCGGCGCTTGCCTTCGCTTCCCATCTGTCCACCTCTCGCATGTAC 195
Db 148 ATGGCCACGGTGGCGCCCTTGCATTCCTTTTGACATACATCTACAGTGTTCAAATAT 207
Qy 196 GGCCTCAAGCCCTTCATCAAGTACGTGTCCGGCATCCCGGACTACTTCAGAGTCCCTC 255
Db 208 GGAATTCGATGCTTTTACTCGGTATCTACCAAGTATGCCCGACTATTTCAAACAAGCATTT 267
Qy 256 CCGGAGGGCTTCACCTGGGAGCGCACCACTACGAGGCGGGGCTTCCTGACCGCC 315
Db 268 CTGACGGAATGATATGAAAGGACTTTTACCTATGAGATGAGAGAGTTGCTACAGCC 327
Qy 316 CACAGGACACCTCCCTCGGACGGCGAGTCCCTGGTGTACAGGTGAAGATCTCGGCAAC 375

Db 328 AGTGGGAATAGCCTTAAGGCNACTGCTTTGAGCAAAATCCAGTTCATGGAGTG 387
 QY 376 AACTTCCCGCGGAGCGCCCGCTGATGAGAACAGCGCCGCTGGAGCCCTCCACC 435
 Db 388 AACTTTCCTGTGATGGACCTGTGATGCGGAAGAAGACAACTGGTTGGAGACCACTTTT 447
 QY 436 GAGATCGTGTAGGAGTGCAGCGGCTGCTGCGCGCCAGTCCCTGATGGCCCTGGAGTC 495
 Db 448 GAGAAAATGACTGCTCGGATGGAAATATTGAAGGTTGATGTCACCGGCTTCCTCATCTG 507
 QY 496 CCGCGGCTGCGCACTGACCTGCGCACCTGCGCACCACTACCGCTCCAGAAAGCCCGCC 555
 Db 508 CAAGGAGTGGCAATTACAGATGCCAATTCACACTTCTTACAGACAAAACCGG-- 565
 QY 556 TCCGCGCTGAAGATCCCGGCTTTCACCTTCAGAGACACCGCATCGATCTGGAGGAG 615
 Db 566 -----TGAGATGCCCAAAACCATGTGTGGAACATCGCATTCGAGGACCGACCTT 618
 QY 616 GTGGAGAGGGCAAGTGTCTAAGCAGTACAGAGCGCGCGTGG 658
 Db 619 GACAAGGTGCAACAGTGTTCAGCTGACGAGCACGCTGTTG 661

RESULT 6

US-08-532-390-40
 ; Sequence 40, Application US/08532390
 ; Patent No. 5795737
 ; GENERAL INFORMATION:
 ; APPLICANT: SEED, BRIAN
 ; APPLICANT: HAAS, JURGEN
 ; TITLE OF INVENTION: High Level Expression of Proteins
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30B
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/532,390
 ; FILING DATE: 22-SEP-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/324,243
 ; FILING DATE: 19-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: LECH, KAREN F.
 ; REGISTRATION NUMBER: 35,238
 ; REFERENCE/DOCKET NUMBER: 00786/294001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 542-5070
 ; TELEFAX: (617) 542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 40:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 762 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-532-390-40

Query Match 16.3%; Score 115; DB 1; Length 762;
 Best Local Similarity 51.2%; Pred. No. 1.5e-14;
 Matches 352; Conservative 0; Mismatches 320; Indels 15; Gaps 3;
 QY 13 CTGCTGACCGAGACCTGCTCCCTTCAGGACCAACCATCATCGAGGCGCCTGTAACGGCCACTAC 72

Db 49 CTGTTCAACCGGGTGGTGCCTCATCTGCTGAGCTGAGCGGACGTAACGGCCACAG 108
 QY 73 TTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTTCBAGGGCACCCAGAGATGAAGATC 132
 Db 109 TTCAAGCTGTCCGGCGAGGGCGAGGCGATGCACCTACGGCAAGCTGACCTGAAGTTC 168
 QY 133 GAGGTGATCGAGGGCGGGCCCTGCGCTTCGACCTTCACATCTCTCCACCTCCCTGCGATG 192
 Db 169 ---ATCTGACCAACCGGCAAGTGCCTGCGCTTCGCGCCACCTCGTGACCACTTCAGC 225
 QY 193 TAGGGTCCAAAGGCTTTCATCAAGTACGTGTCCGGCAT-----CCCCGACTACTTCAAG 246
 Db 226 TAGCGGTGACGTGCTTCAGCGCTACCCGACACATGAAGCAGCAGCACTTCTTCAAG 285
 QY 247 CAGTCCCTCCCGAGGGCTTTCACCTGGAGAGCGCACACACCTAGAGAGACGGGGCTTC 306
 Db 286 TCCGCGCATGCCGAAAGGTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAAC 345
 QY 307 CTGACCGCCACAGCAGCACACTCCCTGAGAGCGGCACTGCTGCTGTACAAGTGAAGATC 366
 Db 346 TACAAGACCGCGCGGAGGTGAAGTTCGAGGGCGACACCTGCTGTAACCGCATCGAGCTG 405
 QY 367 CTGGGCAACAACTTCCCGCGGACCGGCGGCTGTGATGTCAGAAACAAGGCGCGGCTGGGAG 426
 Db 406 AAGGGCATCGACTTCAAGGAGGACGGCAACATCTCTGGGGCACAAAGCTGGAGTACAACCTAC 465
 QY 427 CCTCCACCGAGATCGTGTAGAGGTGAGCGGCTGCGCGGCGCACTCCCTGATGCGC 486
 Db 466 AACAGCCACACGTCTATATCATGSCCAGCAAGCAGAGAGAGCGGATCAAGGTGACTTC 525
 QY 487 CTGGAGTGCCTCCGCGGCTGCGCACCTGACCTGCCACCTGCCACACACCTACCGCTCCAAG 546
 Db 526 AAGATCCGCGCAC-----AACATCGAGGACGGCAGCTGCGAGCTGCGCGACCACTACCGAG 579
 QY 547 AAGCCGCGCTCCGCTGAAAGTGCCTGAGTGCCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 606
 Db 580 CAGAACACCCCATCGGCGAGCGGCGGCTGCTGCTGCCGCAACCACTACTGAGGACCC 639
 QY 607 CTGGAGGAGGTGGAGAGGGCAAGTGTCTACAAGCAGTACGAGGCGCGGCTGGGCGGCTAC 666
 Db 640 CAGTCCGCGCTGAGCAAGACCCCAACGAGAGGCGGATCACATGCTCTCTGAGGATTC 699
 QY 667 TCGAGCGCGCGGCTCCCAAGTGGGC 693
 Db 700 GTGACCGCGCGGATCACTCACGGC 726

RESULT 7

US-08-717-294-40
 ; Sequence 40, Application US/08717294
 ; Patent No. 6114148
 ; GENERAL INFORMATION:
 ; APPLICANT: SEED, BRIAN
 ; APPLICANT: HAAS, JURGEN
 ; TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF
 ; PROTEINS
 ; NUMBER OF SEQUENCES: 110
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Clark & Elbing LLP
 ; STREET: 176 Federal Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/717,294
 ; FILING DATE: 20-SEP-1996

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/345001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 762 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other
US-08-717-294-40

Query Match 16.3%; Score 115; DB 3; Length 762;
Best Local Similarity 51.2%; Pred. No. 1.5e-14;
Matches 352; Conservative 0; Mismatches 320; Indels 15; Gaps 3;

13 CTGCTGACGAGACCATGCCCTTCAGGACACCATCGAGGCGACCGTGAACGCCACTAC 72
19 CTGTTACCGGGTGTGCCATCCTGGTGAAGTGGAGCGGACGTGAACGCCCAAG 108
73 TTCAAGTGACCGGACGAGGCGGAGGCAACCCCTCGAGGCGACCCAGGAGATGAAGATC 132
109 TTCAAGCGTGTCCGCGAGGCGGAGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTC 168
133 GAGGTGATGAGGGGCGGCGCCCTTCCGCTTCCATCATCTGTCTCCACTCTCTGCTG 192
169 ---ATCTGACACCGGCGAGGCTGCCCTGCGCCACCCCTGCTGACCACTTTGAG 225
193 TAGGGTCCAGGCTTCATCAAGTAGTGTCTCGGAT-----CCCGACTACTTCAAG 246
226 TAGGGGTGAGTGTCTTACCGGCTACCCGACCAATGAGGAGGACGAGTCTTCTCAAG 285
247 CAGTTCCTCCCGAGGCTTCACTGAGGAGCGGACCACTTCAAGGAGCGGCGCTTC 306
286 TCCGCCATGCCGAGGCTAGTCCAGGAGCGGACCATCTTCTCAAGGAGCGGCAAC 345
307 CTGACCGCCACAGGACACCTCCCTGGAGCGGAGTCCCTGTGTGTACAGGTGAAGATC 366
346 TACAAGACCGCGCGAGGTAAAGTTCGAGGGGAGACACCTGTGTGAACCGCATCGAGCTG 405
367 CTGGGCAACAACTTCCCGCGGAGCGGCGCTGTATGAGAAACAGGCGGCGCTGGAG 426
406 AAGGGCATGACTTCAAGGAGGAGCGGCAACATCTTGGGCGACAAAGCTGGAGTACAATAC 465
427 CCCTCCACGAGATCGTGTACGAGGTGGAAGGCGTGTGCGGCGGCGGCTCCCTGATGGCC 486
466 AACAGCCACAACTGTATATCATGCGGCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 525
487 CTGAGGTGCGCGCGGTGCGGCTGACCTGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 546
526 AAGATCGCCGAC-----AATCGAGAGCGGAGCGTGTGCTGCGGCGGCGGCGGCGGCGG 579
547 AAGCGCGGCTCCGCGCTGAGATGCGCGGCTTCACTTCGAGGAGGAGGAGGAGGAGGAGGAG 606
580 CAGAACACCCCATCGGCGAGCGGCGGCGTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 639
607 CTGAGGAGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 666
640 CAGTCCGCGCTGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 699
667 TGCAGCGCGCGGCGGCTTCAAGCTGGG 693
700 GTGACCGCGCGGAGTCACTCAGGC 726

RESULT 8
US-09-486-241-31
Sequence 31, Application US/09486241
Patent No. 6472184
GENERAL INFORMATION:
APPLICANT: Hegemann, Peter
TITLE OF INVENTION: METHOD FOR PRODUCING NUCLEIC ACID
TITLE OF INVENTION: POLYMERS
FILE REFERENCE: 3910/OG706
CURRENT APPLICATION NUMBER: US/09/486,241
CURRENT FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: PCT/EP98/05219
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: DE19736591.4
PRIOR FILING DATE: 1997-08-22
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 31
LENGTH: 717
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Modified gene from Aquorea victorea
US-09-486-241-31

Query Match 16.2%; Score 114.6; DB 4; Length 717;
Best Local Similarity 52.1%; Pred. No. 1.7e-14;
Matches 358; Conservative 0; Mismatches 314; Indels 15; Gaps 4;

13 CTGTGACCGGAGACCATGCCCTTCAGGACACCATCGAGGCGACCGTGAACGCCACTAC 72
19 CTGTTACCGGTGTGTTCCCATCTGTTGAGCTGGAGCGGCGGCGGCGGCGGCGGCGGCGG 78
73 TTCAAGTGCACCGGAGGCGGAGGCGAACCCTTCGAGGCGACCCAGGAGATGAAGATC 132
79 TTCTCCGTCTCCGCGAGGCTGAGGCTGACGCCACCTACGGCAAGCTGACCCCTGAAGTTC 138
133 GAGGTGATCGAGGCGGCGGCGGCTGCGCTTCCCTTCCACATCTCTGTCACCTCTCTGATG 192
139 ATCTGCACAC---CGGCAAGCTGCCCTGCGCCACCCCTGGTCAACCTTACC 195
193 TAGGCTCCAGGCTTCATCAAGTACGTGTCTGGGAT-----CCCGACTACTTCAAG 246
196 TAGGCTGTGCTGTCTTCTCCGCTTACCCCGGACCATGAAGGAGGAGGAGGAGGAGGAGG 255
247 CAGTCCCTCCCGAGGCTTCACCTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 306
256 TCCGCCATGCCGAGGCTTACGTGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 315
307 CTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 366
316 TACAAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 375
367 CTGGGCAACAACTTCCCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
376 AAGGCGCATGACTTCAAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 432
427 CCCTCCACGAGATCGTGTACGAGGTGGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 486
433 TACAACCTCCACAACTGTGTATCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 492
487 CTGGAGTGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 546
493 TTCAAGATCC---GCCACAACTATCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 549
547 AAGCGCGCTCCCGCTGAGATGCGCGGCTTCCACTTCGAGGAGGAGGAGGAGGAGGAGGAGGAG 606
550 CAGAACACCCCATCGGCGATGCGCGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 609
607 CTGGAGGAGGTGAGAGGCGGCAAGTGTCTAAGAGAGTACGAGGCGGCGGCGGCGGCGGCGGCTAC 666
610 CAGTCCGCGCTGTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 669

```
QY 667 TCCGACGCGCGCCCTCCAAAGTGGC 693
Db 670 GTACCGGTGCGGATCAACCACGCG 696

RESULT 9
US-08-818-253-7
; Sequence 7, Application US/08818253
; Patent No. 5998204
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: DETECTION OF ANALYTES
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,253
; FILING DATE: 14-MAR-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/043001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1971 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1....1968
US-08-818-253-7

Query Match 16.0%; Score 113; DB 2; Length 1971;
Best Local Similarity 50.9%; Pred. No. 4.1e-14;
Matches 353; Conservative 0; Mismatches 325; Indels 15; Gaps 3;

QY 13 CTGTGACCGGACCATGCGCTTCAGGACCAACCATCGAGGGCACCGTGAACGGCCACTAC 72
Db 1279 CTGTTACCGGGGTGTGTCCTATCTTGGTCTGAGTGGACGGGACGTTAAACGGCCACAAG 1338
QY 73 TTCAAGTGCACCGCAAGGCGAGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATC 132
Db 1339 TTCAGCGTGTCCGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTC 1398
QY 133 GAGGTGATCGAGGGCGGCGCCCTCGCTTCCATCTGTCACATCTCTGTCCACATCTCTGTGATG 192
Db 1399 ATCTGCACCAAC---CGCAAGCTGCGCGTGGCCCTGCGCCACCTCTGTGACCAACCTGACC 1455
QY 193 TACGGCTCAAGGCGCTTTCATCAAGTACGTGTCCGGCAT-----CCCCGACTACTTCAAG 246
Db 1456 TACGGGTGAGTGTCTTTCAGCGCTACCCCGACCAATGAAGCAGCAGCACTTCTTCAAG 1515

QY 247 CAGTCCCTCCCGAGGGCTTACCTGGGAGCGCACCAACCATAGAGACGGCGCTTC 306
Db 1516 TCCGCTATGCCGAGGCTACGTCAGGAGCGCACATCTTCTTCAAGGACGACGGCAAC 1575
QY 307 CTGACCGCCACACGAGACACCTCCCTGGACGCGCATGCTGCTGTGTACAAGGTGAAGATC 366
Db 1576 TACAAGACCCGCGCGAGGTGAAGTTCGAGGGCGCACACCTGTTGAACCGCATCGAGCTG 1635
QY 367 CTGGGCAACAATTCCTCCGCGGACGCGCCGCTGATGCGAGCAACAGGCGCGCTGGGAG 426
Db 1636 AAGGGCATCGACTTCAAGGAGGACGGCAACATCTCTGGGCGCAACGCTGGAGTACAACTAC 1695
QY 427 CCTCCACCGAGATCGTGTACGAGGTGGACGCGCTGCTGCGGCGCAGTCCCTGTATGGCC 486
Db 1696 AACAGCCACACGCTTATATCATGCGGACAGCAAGAGAACGGCATCAAGGTGAATTC 1755
QY 487 CTGGAGTGGCGCGGCTCGCCACCTGACCTGCGACCTGCGACACCACTACCGTCCAAAG 546
Db 1756 AAGATCCGCCAC-----AACATCGAGACGCGCAGCTGCGCGACCACTTACCAG 1809
QY 547 AAGCCCGCTCCGCGCTGAAGATGCGCGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
Db 1810 CAGAACACCCCATCGGCGACGGCCCGCTGCTGCTGCCGACCAACCACTACTGAGCACC 1869
QY 607 CTGGAGAGGTGGAGAAAGGCAAGTGTCTAAGCAAGTACGAGGCGCGCGCTGGCGCGCTAC 666
Db 1870 CAGTCCGCGCTGAGCAAAAGACCCCAACGAGAGCGCGATCACATGCTCTGCTGAGTTC 1929
QY 667 TCGGACGCGCGCCCTCCAAAGCTGGGCGCACCAAC 699
Db 1930 GTGACCGCGCGGGATCACTCTCGGCAAGGAC 1962

RESULT 10
US-08-818-252-7
; Sequence 7, Application US/08818252B
; Patent No. 6197928
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: DETECTION OF ANALYTES
; FILE REFERENCE: 07257/042001
; CURRENT APPLICATION NUMBER: US/08/818,252B
; CURRENT FILING DATE: 1997-03-14
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1971
; TYPE: DNA
; ORGANISM: Aequorea victoria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1968)
US-08-818-252-7

Query Match 16.0%; Score 113; DB 3; Length 1971;
Best Local Similarity 50.9%; Pred. No. 4.1e-14;
Matches 353; Conservative 0; Mismatches 325; Indels 15; Gaps 3;

QY 13 CTGTGACCGGACCATGCGCTTCAGGACCAACCATCGAGGGCACCGTGAACGGCCACTAC 72
Db 1279 CTGTTACCGGGGTGTGTCCTATCTTGGTCTGAGTGGACGGGACGTTAAACGGCCACAAG 1338
QY 73 TTCAAGTGCACCGCAAGGCGAGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATC 132
Db 1339 TTCAGCGTGTCCGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTC 1398
QY 133 GAGGTGATCGAGGGCGGCGCCCTCGCTTCCATCTGTCACATCTCTGTCCACATCTCTGTGATG 192
Db 1399 ATCTGCACCAAC---CGCAAGCTGCGCGTGGCCCTGCGCCACCTCTGTGACCAACCTGACC 1455
QY 193 TACGGCTCAAGGCGCTTTCATCAAGTACGTGTCCGGCAT-----CCCCGACTACTTCAAG 246
Db 1456 TACGGGTGAGTGTCTTTCAGCGCTACCCCGACCAATGAAGCAGCAGCACTTCTTCAAG 1515
```

Db 1456 TACGGCGTGAAGTCTTACCGCGTACCCCGACCATGAAAGCAGCAGCTCTTCAAG 1515
Qy 247 CAGTCCCTCCCGAGGCTTACCTGGAGCGCACACCACTACGAGGACGCGGCTTC 306
Db 1516 TCCGCCATGCCCAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGCAAC 1575
Qy 307 CTGACCGCCACACAGGACACCTCCCTGGAGCGGAGCTGCTGGTGTACAAAGGTGAAGATC 366
Db 1576 TACAGACCCGCGCGAGGTGAAGTTCAGGGGCGACACCTGGTGAACCCATCGAGCTG 1535
Qy 367 CTGGGCAACAATTTCCCGCGAGCGGCCCGCTGATGAGAACAGGCGCGCCCTGGGAG 426
Db 1636 AAGGCACTGACTTCAAGGAGGACGCGCAACATCTCGGGGCAAGCTGGAGTCAACTAC 1695
Qy 427 CCTCCACCGAGATCGTGTACGAGGTGAGCGGCGTGTGCGCGGCCAGTCCCTGATGGCC 486
Db 1696 AACAGCCACACGCTTATATATATGATGCGCGCAAGCAGAGAACGCGCATCAAGGTGAATTC 1755
Qy 487 CTGAGTGTCCCGCGCGTCCGACCTGACCTGACACACCACTACCGCTCCCAAG 546
Db 1756 AAGATCCGCGAC-----AACATCGAGGACGCGAGCTGAGCTCGCGACCACTACCG 1809
Qy 547 AAGCCCGCTCCGCGCTGAAGATGCGCGGCTTCCATTCGAGGACACCGCATCGAGATC 606
Db 1810 CAGAACACCCCATCGCGGACGCGCGCGTGTGCTGCCCGACACCACTACCTGAGCACC 1869
Qy 607 CTGAGGAGGTGGAGAGGCGCAAGTGTACAGCAGTACGAGCGCGCGTGGCGCGCTAC 666
Db 1870 CAGTCCGCTTACGAAAGACCCCAACGAGAGGCGGATCATGTGCTGCTGGAGTTC 1929
Qy 667 TGGAGCGCGCGCGCTCCCAAGTGGGCGCAAC 699
Db 1930 GTGACCGCGCGGATCACTCTCGGCAAGGAC 1962

RESULT 11

US-09-277-716-30
; Sequence 30, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; EARLIER FILING DATE: 1999-03-26
; EARLIER FILING DATE: 1998-10-01
; EARLIER FILING DATE: 1998-10-01
; EARLIER FILING DATE: 1998-06-15
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; FEATURE:
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-277-716-30

Query Match 15.9%; Score 112.4; DB 3; Length 1104;
Best Local Similarity 50.8%; Pred. No. 5e-14;
Matches 269; Conservative 0; Mismatches 261; Indels 0; Gaps 0;
Qy 16 CTGACCGAGACCATGCGCTTCAGGACCAACCATCGAGGCGACCGTGAACGCCACTACTTC 75
Db 64 CTGAAGAGATATATGTCGGCAAAAGCTAGCGTTGAGGAATCTGTGAACATCACTCGTTTTT 123

Qy 76 AAGTGCACCGGCAAGGCGAGGGCAACCCCTCGAGGGCACCCAGGAGATGAAGATCGAG 135
Db 124 TCCATGGAAGGATTTGSAAGGCAATGTATTTATTTGAAACCAATGATGCAATTCGGG 183
Qy 136 GTGATCGAGGGGCGGCCCTCGCTTCGCTTCCACATCCTGTCCACCTCCTGCAATGTAC 195
Db 184 GTTACAAAGGAGGTCCGTTGCCATTCGCTTTCGATATTTGTTCCATAGCTTCCAAATAC 243
Qy 196 GGCTCAAAGGCGCTTCATCAAGTACGTGTCGGGATCCCGGACTACTTCAAGCAGTCCCTC 255
Db 244 GGGAAATCGCACATTTTACGAAATACCCAGACGACATTTGGGGACTACTTTGTTCAATCATTC 303
Qy 256 CCGGAGGGCTTCACCTGGGAGCGCACCACTACGAGGACGCGGCTTCTCGACCGCC 315
Db 304 CCGGCTGATTTTCTACGAAGAATCTACGCTTTGAAGTGGCGCATGTTGCATTC 363
Qy 316 CACCAAGGACACCTCCCTGGAGCGGACCTGCTGTGTGTTCAAGGTGAAGATCTCGGGCAAC 375
Db 364 CGTTCAGATATTAAGTTTAGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGGCAAC 423
Qy 376 AACTTCCCGCGAGCGCGCGTGTGATGAGACAGGCGCGCGCTGGGAGCCCTCCACC 435
Db 424 GGTTCCTTCTAGTAACGAGCCCGTGTGATGCAAAAGCCATCCTCGGATGGAGCCATCGTTT 483
Qy 436 GAGATCGTGTACGAGGTGGACGCGCTGCTGCGCGCCAGTCCCTGATGCGCCCTGGAGTGC 495
Db 484 GAGGTGGTCTACATCAACAGCGCGCTTCTGGTGGCGAAGTAGATCTCGTTTACAACTC 543
Qy 496 CCGGCGGTGCGCACCTGACCTGCCACCTGCACACCACTACCGCTCCAA 545
Db 544 GAGTCAGGGAACATATTACTCGTGCACATGAAGAGCTTTTACAGATCAA 593

RESULT 12

US-09-609-161B-30
; Sequence 30, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; PRIOR FILING DATE: 2000-06-30
; PRIOR FILING DATE: 1999-03-26
; PRIOR FILING DATE: 1998-10-01
; PRIOR FILING DATE: 1998-10-01
; PRIOR FILING DATE: 1998-06-15
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-609-161B-30

Query Match 15.9%; Score 112.4; DB 4; Length 1104;
Best Local Similarity 50.8%; Pred. No. 5e-14;
Matches 269; Conservative 0; Mismatches 261; Indels 0; Gaps 0;
Qy 16 CTGACCGAGACCATGCGCTTCAGGACCAACCATCGAGGCGACCGTGAACGCCACTACTTC 75
Db 64 CTGAAGAGATATATGTCGGCAAAAGCTAGCGTTGAGGAATCTGTGAACATCACTCGTTTTT 123

Db 64 CTGAAGAGATTATGTGGGCAAAAGCTAGCGTTTGAAGGAATCGTGAACAATCACGTTTTT 123
Qy 76 AAGTGCACCGCAAGGCGGCAACCCCTCGAGGCGCACCCAGGAGATGAAGATCGAG 135
Db 124 TCCATGAGAGATTGGGAAAGGCAATGTATTTTGGAAACCAATTTGATGCAATCCGG 183
Qy 136 GTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCTCTGTCCACCTCTCGATGTAC 195
Db 184 GTTACAAAGGAGGTCCGTTGCCATTCGCTTTCGATATTTGTTCCATAGCTTTCCATAC 243
Qy 196 GSCTCCAGGCGCTTCATCAAGTAGTGTGCGGCATCCCGACTACTTCAAGAGAGTCCCTC 255
Db 244 GGGATCCGCTTTCACGAATATCCAGACGACATTTGGGACTACTTTGTTTCATCATTC 303
Qy 256 CCGAGGCGCTTCACCTGGAGCGCACCAACCTACAGAGAGCGCGCTTCCTGACCGCC 315
Db 304 CCGGTGGATTTTCTACGAAGAAATCTACGCTTTGAAGATGCGGCAATTTGTGACAT 363
Qy 316 CACAGGACACTCCCTGGAGCGGACTGCTGCTGTACAGGTGAAGATCCTGGGCAAC 375
Db 364 CGTTCAGATATAAGTTTGAAGATGATAGTTCCACTACAAAGTGGAGTATAGAGGCAAC 423
Qy 376 AACTTCCCGCGGCGGCGCTGATGAGAAACAAAGCGCGCGCTCGGAGCGCTCCACCC 435
Db 424 GGTTCCTCTAGTAAGCGGACCGTGTATGCAAAAGCCATCTCTCGCATGGAGCCATCGTT 483
Qy 436 GAGATCGGTGACGAGGTGGAGCGGTGCTGCGGCGGAGTCCCTGTATGGCGCTGGAGTGC 495
Db 484 GAGGTGCTCTATGAAACAGCGGCGTCTGCTGGCGGAAAGTAGATCTCGTTTACAAACTC 543
Qy 496 CCGGCGGTGCGCACTGACCTGACCTGCGACCTGACACCACTACCGCTCCAA 545
Db 544 GAGTCAGGAATATTACTGTCGCACATGAAGGTTTACAGATCCAA 593

RESULT 13

US-09-277-716-31
; Sequence 31, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; PRIOR FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1279
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(720)
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert B)
US-09-277-716-31

Query Match 15.9%; Score 112.4; DB 3; Length 1279;
Best Local Similarity 50.8%; Pred. No. 5,1e-14;
Matches 269; Conservative 0; Mismatches 261; Indels 0; Gaps 0;
Qy 16 CTGACCGAGACCATCCCTTCAGGACCACTCGAGGCGACCGTGAACCGGCACACTATTC 75
Db 37 CTGAAGAGATTATGTGCGCAAAAGCTAGCGTTTGAAGGAATCGTGAACAATCACGTTTTT 96

Qy 76 AAGTGCACCGCAAGGCGGCAACCCCTCGAGGCGCACCCAGGAGATGAAGATCGAG 135
Db 97 TCCATGAGAGATTGGGAAAGGCAATGTATTTTGGAAACCAATTTGATGCAATCCGG 156
Qy 136 GTGATCGAGGCGGCGCCCTTCGCTTCGCTTCACATCTCTGTCCACCTCTCGATGTAC 195
Db 157 GTTACAAAGGAGGTCCGTTGCCATTCGCTTTCGACATTTGTTCCATAGCTTTCCATAC 216
Qy 196 GSCTCCAGGCGCTTCATCAAGTAGTGTGCGGCATCCCGACTACTTCAAGAGAGTCCCTC 255
Db 217 GGGAAATCGCACTTTTACGAATATCCAGACGACATTTGGGACTACTTTGTTTCATCATTT 276
Qy 256 CCGAGGCGCTTCACCTGGAGCGCACCAACCTACAGAGAGCGCGCTTCCTGACCGCC 315
Db 277 CCGGTGGATTTTCTACGAAGAAATCTACGCTTTGAAGATGCGGCAATTTGTGACATT 336
Qy 316 CACAGGACACTCCCTGGAGCGGACTGCTGCTGTACAGGTGAAGATCCTGGGCAAC 375
Db 337 CGTTCAGATATAAGTTTGAAGATGATAGTTCCACTACAAAGTGGAGTATAGAGGCAAC 396
Qy 376 AACTTCCCGCGGCGGCGCTGATGAGAAACAAAGCGCGCGCTGGAGCGCTCCACCC 435
Db 397 GGTTCCTCTAGTAAGCGGACCGTGTATGCAAAAGCCATCTCTCGCATGGAGCCATCGTT 456
Qy 436 GAGATCGGTGACGAGGTGGAGCGGTGCTGCGGCGGAGTCCCTGTATGGCGCTGGAGTGC 495
Db 457 GAGGTGCTCTATGAAACAGCGGCGTCTGCTGGCGGAAAGTAGATCTCGTTTACAAACTC 516
Qy 496 CCGGCGGTGCGCACTGACCTGACCTGCGACCTGACACCACTACCGCTCCAA 545
Db 517 GAGTCAGGAATATTACTGTCGCACATGAAGGTTTACAGATCCAA 566

RESULT 14

US-09-609-161B-31
; Sequence 31, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1279
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(720)
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert B)
US-09-609-161B-31

Query Match 15.9%; Score 112.4; DB 4; Length 1279;
Best Local Similarity 50.8%; Pred. No. 5,1e-14;
Matches 269; Conservative 0; Mismatches 261; Indels 0; Gaps 0;
Qy 16 CTGACCGAGACCATCCCTTCAGGACCACTCGAGGCGACCGTGAACCGGCACACTATTC 75


```
Db 37 CTGAAGAGATATGTCGCAAAAGCTAGCGTTGAAGGAATCGTGAACAAATCAGCTTTT 96
Qy 76 AAGTGCACCGGCAAGGCGGAGGCGCAACCCCTCGAGGCGACCCAGGAGATGAAGATCGAG 135
Db 97 TCCATGGAAGGATTTGGAAGAGCAATGATATTTATTTGGAAACCAATTTGATGCAATCCGG 156
Qy 136 GTGATCGAGGGCGGCGCCCTCGCCTTCGCTTCACATCCTGTCCACCTCCTGCAATGATC 195
Db 157 GTTACAAAGGAGGTCGCTTGCCTTCGCTTCGCAATTTGTTCCATAGCTTTTCCAATAC 216
Qy 196 GGCTCCAAAGGCCCTTCATCAAGTAGTACGTGTCGGGATCCCGGACTTCAAGCAGTCCCTC 255
Db 217 GGAATTCGCACTTCACGAATATCCAGACGACATTCGCGACTACTTTGTTCAATCATTT 276
Qy 256 CCGAGGGCTTACCTGAGGAGCGCACCACTACGAGAGCGGCGCTTCCTGACCGCC 315
Db 277 CCGGCTGGATTTTACGAAAGAAATCTACGCTTTGAAGATGGCGCCATTGTTGACATT 336
Qy 316 CACGAGACACCTCCCTGAGCGGAGCTGCTGCTGTACAAAGTGAAGATCCTGGGCAAC 375
Db 337 CGTTCAGATATAGTTTAGAGATGATAGTTCCACTACAAAGTGAAGTATAGAGGCAAC 396
Qy 376 AACTTCCCGCCGACCGCCCGTGTGATGCAAGAACAGGCGGCGCTGGAGCCCTCCACC 435
Db 397 GGTTCCTAGTAAACGACCGGTGATGCAAAAGCCATCCTCGCATGGAGCCATCGTTT 456
Qy 436 GAGATCGTACGAGGTGGAGCGGTGCTGCGGCGGCGAGTCCCTGTATGCGCCCTGGAGTGC 495
Db 457 GAGGTGCTACATATGAACAGCGCGGTTCGTGGGCGAGTAGATCTCGTTTACAAATC 516
Qy 496 CCGGGGGGTGCGCACCTGACCTCCACCTGCACACACCTTACCGTCCAA 545
Db 517 GAGTCAGGAACTATTACTCGTCCACATGAANAACGTTTACAGATCCAA 566
```

RESULT 15

```
US-09-513-783A-45
; Sequence 45, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 45
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(717)
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-45
```

```
Query Match 15.8%; Score 111.8; DB 4; Length 717;
Best Local Similarity 50.9%; Pred. No. 6.2e-14;
Matches 350; Conservative 0; Mismatches 322; Indels 15; Gaps 3;

Qy 13 CTGCTGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGCCGTTGAACGGCCACTAC 72
Db 22 CTGTTACCGGGGTGTGCCCATCCTGTCGAGTGGACGCGGACGTAAACGGCCCAAG 81
Qy 73 TTCAAGTGCACCGGAGGCGGCAACCCCTCGAGGCGACCCAGGAGATGAAGATC 132
Db 82 TTCAGCGTGTCCGCGGAGGCGAGGCGGATGCGACCTACCGGCAAGCTGACCCCTGAATTC 141
Qy 133 GAGGTGATGAGGCGGCGCCCTTCGCTTCACATCCTGTCCACCTCCTGTCATG 192
```

```
Db 142 ATCTGCACCAC---CGCAAGCTGCCGTGCCCTGGCCACCCTCTGTCGACCAACCTGACC 198
Qy 193 TACGGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCAT-----CCCGGACTACTTCAAG 246
Db 199 TACGGCGTGCAGTGTCTTACGCCGTACCCCGACACATGAAGCAGCAGCATCTTCTTCAAG 258
Qy 247 CAGTCCCTCCCGGAGGCTTCACCTGGGAGCGCACCACTACGAGGACGGCGGCTTC 306
Db 259 TCCGGCATGCCCGAAGGCTAGCTCAGAGGCGACCATCTTCTTCAAGGACGACGGCAAC 318
Qy 307 CTGACCGCCCAACAGGACACCTCCTGGAAGCGGCACTGCTGCTGTGTAACAAGTGAAGATC 366
Db 319 TACAAGACCGCGCGAGGTGAAGTTCGAGGCGACACACCTGTGTGAACCGCATCGAGCTG 378
Qy 367 CTGGCAACAACCTTCCCGCGGAGCGGCGCTGTGATGAGACAAAGCGGCGGCTGGAG 426
Db 379 AAGGCGATCGACTTCAAGGAGGCGGCAACATCTCTGGGCGACAAGCTGGAGTACAATAC 438
Qy 427 CCCTCCACCGGAGATCGTGTACGAGGTGGACGGCGGTGCTGCGCGGCCAAGTCCCTGATGCC 486
Db 439 AACGCCACAACGTTCTATATCATGGCGGACAAGCAGAAGAACGGCATCAAGGTGAACCTTC 498
Qy 487 CTGAGTGCCTCCCGGGGTGCGCACCTGACCTGCCACCTGCAACACCACTACCTGAGCACC 546
Db 499 AAGATCCGCCAC-----AACATCGAGACGCGAGCTGCGGCACTACTACCG 552
Qy 547 AAGCGCGCTCCGCGCTGAAGATGCCCGGCTTCCACTTCGAGGACCAACCGCATCGAGATC 606
Db 553 CAGAACACCCCATCGGCGACGGCGCTGTGCTGCGCCGACACCACTACTCTGAGCACC 612
Qy 607 CTGGAGGAGGTGGAGAGGCGCAAGTGTCTAAGCAGTACGAGGCGCGCTGGCGGCTAC 666
Db 613 CAGTCCGCGCTGAGCAAAAGACCCCAACGAGAAGCGGATCATGTTCTCTCTGAGTTC 672
Qy 667 TCGGACGCGCGCCCTCCCAAGCTGGC 693
Db 673 GTGACCGCGCGGGATCCTCTCGGC 699
```

Search completed: August 3, 2004, 15:50:33
Job time : 73.4657 secs

Blank Sheet USPTO

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 14:04:25 : Search time 514.043 Seconds
(without alignments)
6743.643 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 707.

Sequence: 1 ggatcgctcctcgtgac.....ctggcccaactgaagctt 707

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	707	100.0	707	14	US-10-006-922-39
2	707	100.0	707	15	US-10-081-864-21
3	678.4	96.0	699	14	US-10-006-922-41
4	669	94.6	696	17	US-10-724-178-17
5	649.2	91.8	634	15	US-10-081-864-23
6	571.4	80.8	591	17	US-10-724-178-1046
7	532.6	75.3	555	17	US-10-724-178-1064
8	481.6	68.1	504	17	US-10-724-178-1060
9	440.2	62.3	696	14	US-10-006-922-13
10	440.2	62.3	696	15	US-10-081-864-9
11	438.4	62.0	456	17	US-10-724-178-1052
12	436.8	61.8	767	15	US-10-155-809-1
13	435.2	61.6	767	15	US-10-155-809-3
14	433.6	61.3	767	15	US-10-155-809-5

15	417.8	59.1	687	9	US-09-976-673-11	Sequence 11, Appl
16	417.8	59.1	687	9	US-09-976-673-13	Sequence 13, Appl
17	417.8	59.1	1396	9	US-09-976-673-15	Sequence 15, Appl
18	417.8	59.1	1424	9	US-09-976-673-17	Sequence 17, Appl
19	387.6	54.8	405	17	US-10-724-178-1050	Sequence 1050, Ap
20	345.6	48.9	363	17	US-10-724-178-1058	Sequence 1058, Ap
21	323.4	45.7	333	17	US-10-724-178-1056	Sequence 1056, Ap
22	319.6	45.2	649	17	US-10-439-262-12	Sequence 12, Appl
23	315	44.6	675	14	US-10-060-922-38	Sequence 38, Appl
24	315	44.6	675	15	US-10-081-864-13	Sequence 13, Appl
25	315	44.6	1050	14	US-10-060-857-7	Sequence 7, Appl
26	313.4	44.3	678	14	US-10-006-922-36	Sequence 36, Appl
27	313.4	44.3	678	15	US-10-081-864-14	Sequence 14, Appl
28	313.4	44.3	678	15	US-10-315-920-5	Sequence 5, Appl
29	313.4	44.3	681	15	US-10-121-258-7	Sequence 7, Appl
30	312.6	44.2	675	17	US-10-724-178-15	Sequence 15, Appl
31	312.6	44.2	678	15	US-10-121-258-9	Sequence 9, Appl
32	310.2	43.9	678	15	US-10-081-864-7	Sequence 7, Appl
33	310.2	43.9	678	15	US-10-121-258-5	Sequence 5, Appl
34	310.2	43.9	678	15	US-10-315-920-1	Sequence 1, Appl
35	308.6	43.6	678	15	US-10-315-920-3	Sequence 3, Appl
36	308.6	43.6	681	15	US-10-121-258-3	Sequence 23, Appl
37	308.6	43.6	681	15	US-10-121-258-23	Sequence 51, Appl
38	308.6	43.6	1638	15	US-10-214-932-51	Sequence 29, Appl
39	308.6	43.6	1647	15	US-10-214-932-75	Sequence 16, Appl
40	308.6	43.6	4692	15	US-10-161-403-29	Sequence 45, Appl
41	308.6	43.6	4692	17	US-10-433-640-16	Sequence 35, Appl
42	308.6	43.6	6984	14	US-10-001-189-45	Sequence 37, Appl
43	307	43.4	681	14	US-10-006-922-37	Sequence 1, Appl
44	307	43.4	681	14	US-10-006-922-37	
45	307	43.4	723	15	US-10-152-296-1	

ALIGNMENTS

RESULT 1

US-10-006-922-39
Sequence 39, Application US/10006922
Publication No. US20020197676A1
GENERAL INFORMATION:
APPLICANT: Lukyanov, Sergey A
APPLICANT: Fradkov, Arcady F.
APPLICANT: Labas, Yulii A.
APPLICANT: Matz, Mikhail V.
APPLICANT: Tersikh, Alexey
TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
FILE REFERENCE: CLON-035CIP
CURRENT APPLICATION NUMBER: US/10/006,922
CURRENT FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 09/120,330
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/457,898
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,144
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/458,477
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/457,556
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 09/444,338
PRIOR FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39
LENGTH: 707
TYPE: DNA
ORGANISM: Anemonia sulcata
US-10-006-922-39

Query Match 100.0% Score 707; DB 14; Length 707;
Best Local Similarity 100.0%; Pred. No. 1e-159;

Matches 707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATCCGCTCCCTGCTGACGAGACCATGCTTCAAGGACCACTCGAGGGACCCGTG 60
DB 1 GGATCCGCTCCCTGCTGACGAGACCATGCTTCAAGGACCACTCGAGGGACCCGTG 60

QY 61 AACGGCCACTACTCAAGTGTACCGGCAAGGCGAGGGCAACCCCTCGAGGGACCCAG 120
DB 61 AACGGCCACTACTCAAGTGTACCGGCAAGGCGAGGGCAACCCCTCGAGGGACCCAG 120

QY 121 GAGATGAAGTGTACGAGGGCGGCGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 180
DB 121 GAGATGAAGTGTACGAGGGCGGCGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 180

QY 181 ACCTCTGCTGATGTACGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 240
DB 181 ACCTCTGCTGATGTACGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 240

QY 241 TTCAAGCATGCTTCCCGGAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 300
DB 241 TTCAAGCATGCTTCCCGGAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 300

QY 301 GGCTTCTGAGCGGCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 360
DB 301 GGCTTCTGAGCGGCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 360

QY 361 AAGATCTGAGCGGCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 420
DB 361 AAGATCTGAGCGGCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 420

QY 421 TGGGAGCGGCTTCAAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 480
DB 421 TGGGAGCGGCTTCAAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 480

QY 481 ATGGCGCTGAGTGTCCCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 540
DB 481 ATGGCGCTGAGTGTCCCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 540

QY 541 TCCAGAGCGGCGGCGGCTTCAAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 600
DB 541 TCCAGAGCGGCGGCGGCTTCAAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 600

QY 601 GAGATCTGAGGAGTGTGAGAGGCAAGTGTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 660
DB 601 GAGATCTGAGGAGTGTGAGAGGCAAGTGTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 660

QY 661 CGCTACTGAGCGGCGGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 707
DB 661 CGCTACTGAGCGGCGGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 707

RESULT 2

US-10-081-864-21
; Sequence 21, Application US/10081864
; Publication No. US2003002287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexander
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US2003002287A1 Aggregating Fluorescent Proteins and
; FILE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/005,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21

Handwritten signature

LENGTH: 707
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: non-aggregating mutant
US-10-081-864-21

Query Match 100.0%; Score 707; DB 15; Length 707;

Best Local Similarity 100.0%; Pred. No. 1e-159; Mismatches 0; Indels 0; Gaps 0;
Matches 707; Conservative 0;

QY 1 GGATCCGCTCCCTGCTGACGAGACCATGCTTCAAGGACCACTCGAGGGACCCGTG 60
DB 1 GGATCCGCTCCCTGCTGACGAGACCATGCTTCAAGGACCACTCGAGGGACCCGTG 60

QY 61 AACGGCCACTACTCAAGTGTACCGGCAAGGCGAGGGCAACCCCTCGAGGGACCCAG 120
DB 61 AACGGCCACTACTCAAGTGTACCGGCAAGGCGAGGGCAACCCCTCGAGGGACCCAG 120

QY 121 GAGATGAAGTGTACGAGGGCGGCGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 180
DB 121 GAGATGAAGTGTACGAGGGCGGCGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 180

QY 181 ACCTCTGCTGATGTACGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 240
DB 181 ACCTCTGCTGATGTACGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 240

QY 241 TTCAAGCATGCTTCCCGGAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 300
DB 241 TTCAAGCATGCTTCCCGGAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 300

QY 301 GGCTTCTGAGCGGCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 360
DB 301 GGCTTCTGAGCGGCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 360

QY 361 AAGATCTGAGCGGCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 420
DB 361 AAGATCTGAGCGGCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 420

QY 421 TGGGAGCGGCTTCAAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 480
DB 421 TGGGAGCGGCTTCAAGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 480

QY 481 ATGGCGCTGAGTGTCCCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 540
DB 481 ATGGCGCTGAGTGTCCCGGCGGCTTCAAGGCTTCAAGTGTACCGGCAACCCCTCGAGGGACCCAG 540

QY 541 TCCAGAGCGGCGGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 600
DB 541 TCCAGAGCGGCGGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 600

QY 601 GAGATCTGAGGAGTGTGAGAGGCAAGTGTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 660
DB 601 GAGATCTGAGGAGTGTGAGAGGCAAGTGTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 660

QY 661 CGCTACTGAGCGGCGGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 707
DB 661 CGCTACTGAGCGGCGGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTT 707

RESULT 3

US-10-006-922-41
; Sequence 41, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1 Chromophores/Fluorophores and
; FILE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP

```

; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Anemonia sulcata
US-10-006-922-41

```

Query Match	96.0%;	Score 678.4;	DB 14;	Length 699;
Best Local Similarity	98.4%;	Pred. No. 7.2e-153;		
Matches 685;	Conservative 0;	Mismatches 11;	Indels 0;	Gaps 0;
QY	7	GCTCCCTGCTCAGCGAGACCATGCCTTCAGGAGCACCATTCAGAGGCGACCGTGAACGCG	66	
DB	4	GCTCTCTCTGAGAGAGCAATGCGCTTCAAGACCAACCATCGAGGCGACCGTGAACGCG	63	
QY	67	CACCTACTCAAGTGCAACCGGCAAGGGGAGGCGCAACCCCTCGAGGGCAACCCAGGAGATG	126	
DB	64	CACCTACTCAAGTGCAACCGGCAAGGGGAGGCGCAACCCCTTCAGGGGCAACCCAGGAGATG	123	
QY	127	AAGATCGAGGTGATCAGAGGCGGCCCTCGCCCTTGCCTTCGCATCTCTGTCCACCTCC	186	
DB	124	AAGATCGAGGTGATCAGAGGCGGCCCTCGCCCTTGCCTTCGCATCTCTGTCCACCTCC	183	
QY	187	TGATGTACGGTCCAAAGGCTTCATCAAGTAAGTTCGGGCATCCCGACTACTTCAAG	246	
DB	184	TGATGTACGGTCCAAAGGCTTCATCAAGTAAGTTCGGGCATCCCGACTACTTCAAG	243	
QY	247	CAGTCCCTCCCGAGGGCTTCACCTGGGAGGCGACCAACCACTTACGAGGAGCGCGGCTTC	306	
DB	244	CAGTCTCTCCCGAGGGCTTCACCTGGGAGGCGACCAACCACTTACGAGGAGCGCGGCTTC	303	
QY	307	CTGACCGGCCACCAAGCAACCTCCCTGGAACGGGACTGCTGTGTGTAAGGTGAAGATC	366	
DB	304	CTGACCGGCCACCAAGCAACCTCCCTGGAACGGGACTGCTGTGTGTAAGGTGAAGATC	363	
QY	367	CTGGGCAACAACCTCCCGCGAGCGGCCGCTGATGCAGAACAAAGCGCGCGCTGGGAG	426	
DB	364	CTGGGCAACAACCTCCCGCGAGCGGCCGCTGATGCAGAACAAAGCGCGCGCTGGGAG	423	
QY	427	CCCTCCACGAGATCGTGTACGAGGTGGAAGGGGTGCTGGCGGCGAGTCTCTGTAGTCC	486	
DB	424	CCCTCCACGAGATCGTGTACGAGGTGGAAGGGGTGCTGGCGGCGAGTCTCTGTAGTCC	483	
QY	487	CTGGAGTGCCCGGGGGTGCACCTGACTTGCACTGCACTGCACACCACTACCGCTCCAG	546	
DB	484	CTGAGTGTCCCGGGGGCGGCACTGACTTGCACTGCACTGCACACCACTACCGCTCCAG	543	
QY	547	AAGCCCGCTTCGCGCTGAAGATCCCGGTTCCACTTGAGGACCAACCGCATCGAGATC	606	
DB	544	AAGCCCGCTTCGCGCTGAAGATCCCGGTTCCACTTGAGGACCAACCGCATCGAGATC	603	
QY	607	CTGGAGGAGGTGGAAGGGCAAGTGTCTCAAGCGTACGAGCGCGCTGGGGCGCTAC	666	
DB	604	ATGGAGGAGGTGGAAGGGCAAGTGTCTCAAGCGTACGAGCGCGCTGGGGCGCTAC	663	
QY	667	TGCGAGCGCGCCCTCCAAAGCTGGGCGCAACTGA	702	
DB	664	TGCGAGCGCGCCCTCCAAAGCTGGGCGCAACTGA	699	

RESULT 4

US-10-724-178-17
US-10-724-178-17
Sequence 17, Application US/10724178
Publication No. US20040137528A1
GENERAL INFORMATION:
APPLICANT: Odyssey Thera, Inc.
APPLICANT: Michnick, Stephen
APPLICANT: MacDonald, Marnie
APPLICANT: Lamerdin, James
TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
FILE REFERENCE: ODDY007
CURRENT APPLICATION NUMBER: US/10/724,178
PRIORITY FILING DATE: 2003-12-01
PRIOR APPLICATION NUMBER: US 60/461,133
PRIOR FILING DATE: 2003-04-09
NUMBER OF SEQ ID NOS: 1067
SOFTWARE: PatentIn version 3.0
SEQ ID NO 17
LENGTH: 696
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: "KPPI" a mutated variant of kindling fluorescent prote
OTHER INFORMATION: m A. sulcata
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(696)
US-10-724-178-17

Query Match 94.6%; Score 669; DB 17; Length 696;
Best Local Similarity 97.8%; Pred. No. 1.3e-150;
Matches 678; Conservative 0; Mismatches 15; Indels 0; Gaps 0

QY	7	GCTCTCTGCTGACCGAGACCATGCGCTTCAGGACACCATCGAGGGCACCGTGAACGGC	66
DB	4	GCTCTCTGCTGACCGAGACCATGCGCTTCAGGACACCATCGAGGGCACCGTGAACGGC	63
QY	67	CACCTACTTCAAGTGCACGGCGAGGGCAGGGCAACCCCTCGAGGGCACCCAGAGATG	12
DB	64	CACCTGCTTCAAGTGCATCGGCAAGGGCAGGGCAACCCCTTCAGAGGGCACCCAGAGATG	12
QY	127	AAGATCGAGGTGATCGAGGGCGGCCCTTGCCCTTTCGCCTTCCACATCTCTGCACCTCC	18
DB	124	AAGATCGAGGTGATCGAGGGCGGCCCTTGCCCTTTCGCCTTCCACATCTCTGCACCTCC	18
QY	187	TGCATGTACGGCTTCAAGCGCTTCATCAAGTACGTGTCGGCATGCCGACTCTTCAAG	24
DB	184	TGCATGTACGGCTTCAAGACCTTCATCAAGTACGTGTCGGCATGCCGACTCTTCAAG	24
QY	247	CAGTCCCTCCCGAGGGCTTCACTTGGAGGGCGCACCACTACGAGGACGGCGGCTTC	30
DB	244	CAGTCCCTCCCGAGGGCTTCACTTGGAGGGCGCACCACTACGAGGACGGCGGCTTC	30
QY	307	CTGACGGCCACACGAGACACCTCCCTGACGGCGACTGCTGGTGATCAAGGTGAAGATC	36
DB	304	CTGACGGCCACACGAGACACCTCCCTGACGGCGACTGCTGGTGATCAAGGTGAAGATC	36
QY	367	CTGGGGCAACAATTCCCGCCGACGGCCCTGTGATGAGAAACAAGCCGGCCGTGGAG	42
DB	364	CTGGGTAAACAATTCCCGCCGACGGCCCTGTGATGAGAAACAAGCTGGCCGTGGAG	42
QY	427	CCTCCACCGAGATCGTGTACGAGGTGACGGCGTGTGCGCGGCCACTGCTCATGGCC	48
DB	424	CCCGNACCGAGATCGTGTGACGAGTGAACGGCGTGTGCGCGGCCAGTCCCTCATGGCC	48
QY	487	CTGGAGTGCCTCGGGGTGCGCACTGACTTGCACCTGTGACACCACTACCGTCCAAG	54
DB	484	CTGAAGTGCCTCGGGGGCGGCACCTGACCTGCCACTGCAACCACTACCGTCCAAG	54
QY	547	AAGCCCGCTCGCGCCTGAAGATGCCGGCTTCCACTTCGAGGACCAACCGCATCGATC	60

Db 303 GGTGGCGGCTGGAGCCCGAAACCGAGATCGTGTACAGGTGAGCGGCTGTGGCGG 362
QY 471 CCACTCCCTGATCGCCCTGAGTGCCTGGCGGCTGCCACCTGACCTGCACCTGCACAC 530
Db 363 CCACTCCCTGATGGCCCTGAAAGTGCCTGGCGGCTGCCACCTGACCTGCACAC 422
QY 531 CACCTACCGCTCCAAGAAGCCCTCCCGCTTGAAGATGCGCGGCTTCCACTTCGAGGA 590
Db 423 CACCTACCGCTCCAAGAAGCCCTCCCGCTTGAAGATGCGCGGCTTCCACTTCGAGGA 482
QY 591 CACCGCATCGATCTCGAGAGGTGAGAGGCGAGTGTACAAAGCAGTACAGAGC 550
Db 483 CACCGCATCGATCATGAGAGGTGAGAGGCGAGTGTACAAAGCAGTACAGAGC 542
QY 651 CGCGCTGGCGGCTACTGCGACCGCCCGCTTCAAGCTGGGCGCAAC 699
Db 543 CGCGCTGGCGGCTACTGCGACCGCCCGCTTCAAGCTGGGCGCAAC 591

RESULT 7

US-10-724-178-1064
; Sequence 1064, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1064
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KFP FlE, with position 1 Met removed
; NAME/KEY: CDS
; LOCATION: (1)..(555)
; OTHER INFORMATION: KFP FlE corresponds to aa residues 1-186 of KPPI
US-10-724-178-1064

Query Match 75.3%; Score 532.6; DB 17; Length 555;
Best Local Similarity 97.5%; Pred. No. 5.2e-118;
Matches 541; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 7 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 66
Db 1 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 60
QY 67 CACTACTTCAAGTCACCGGAGGGGAGGCAACCCCTTCGAGGCGACCCAGGAGATG 126
Db 61 CACTGTCTCAAGTGCATCGGCAAGGGGAGGCAACCCCTTCGAGGCGACCCAGGAGATG 120
QY 127 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCGCTTCCACATCTCTGTCCACCTCC 186
Db 121 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCGCTTCCACATCTCTGTCCACCTCC 180
QY 187 TGCATGTACGGCTCCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db 181 TGCATGTACGGCTCCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 240
QY 247 CAGTCCCTCCCGAGGCTTCACCTGGAGCGCACCAACCTACGAGGACGCGGCTTC 306
Db 241 CAGTCCCTCCCGAGGCTTCACCTGGAGCGCACCAACCTACGAGGACGCGGCTTC 300

QY 307 CTGACCGCCACCGAGCACCTCCCTGAGCGGCGACCTGCCTGTGTACAAGTGAAGATC 366
Db 301 CTGACCGCCACCGAGCACCTCCCTGAGCGGCGACCTGCCTGTGTACAAGTGAAGATC 360
QY 367 CTGGCAACAATTTCCCGCGCGAGCGCCCTGTATGAGAACAAAGCGCGGCTGGAG 426
Db 361 CTGGTAAACAATTTCCCGCGCGAGCGCCCTGTATGAGAACAAAGCGCGGCTGGAG 420
QY 427 CCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGTGCGCGGCGAGTCCCTGTATGGCC 486
Db 421 CCGGAACCGAGATCGTGTACGAGGTGAGCGGCTGTGCGCGGCGAGTCCCTGTATGGCC 480
QY 487 CTGAGTCCCGCGCGGCTGCACCTGACCTGCACCTGCAACACCTACCGCTCCAAG 546
Db 481 CTGAAGTCCCGCGCGGCGCACCTGACCTGCCACCTGCACACCTACCGCTCCAAG 540
QY 547 AAGCCCGCTCCGCC 561
Db 541 AAGCCCGCTCCGCC 555

RESULT 8

US-10-724-178-1060
; Sequence 1060, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1060
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KFP FlE, with position 1 Met removed
; NAME/KEY: CDS
; LOCATION: (1)..(504)
; OTHER INFORMATION: KFP FlE corresponds to aa residues 1-169 of KPPI
US-10-724-178-1060

Query Match 68.1%; Score 481.6; DB 17; Length 504;
Best Local Similarity 97.2%; Pred. No. 8.1e-106;
Matches 490; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 7 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 66
Db 1 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCAACCATCGAGGCGACCGTGAACGGC 60
QY 67 CACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTTCGAGGCGACCCAGGAGATG 126
Db 61 CACTGTCTCAAGTGCATCGGCAAGGGCGAGGCAACCCCTTCGAGGCGACCCAGGAGATG 120
QY 127 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCGCTTCCACATCTCTGTCCACCTCC 186
Db 121 AAGATCGAGGTGATCGAGGGCGGCGCCCTGCGCTTCCACATCTCTGTCCACCTCC 180
QY 187 TGCATGTACGGCTCCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db 181 TGCATGTACGGCTCCAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 240
QY 247 CAGTCCCTCCCGAGGCTTCACCTGGAGCGCACCAACCTACGAGGACGCGGCTTC 306

Db 241 CAGTCTTCCCGAGGGCTTCACTGGAGCGCACCACTACGAGGACGGCGGCTTC 300
QY 307 CTGACCGCCACACAGGACACTCCTCGACGGGACTGCTGTGTACAAAGGTCAAGATC 366
Db 301 CTGACCGCCACACAGGACACTCCTCGACGGGACTGCTGTGTACAAAGGTCAAGATC 360
QY 367 CTGGCAACAACTTCCCGCGGACGCGCCCGTGTATGAGAAACAAAGGCGCGCGTGGAG 426
Db 361 CTGGTAACAACTTCCCGCGGACGCGCCCGTGTATGAGAAACAAAGGTCGCGCGTGGAG 420
QY 427 CCTCCACCGAGATGCTGTACAGGTGACGCGGCTGTGCGCGCCAGTCCCTGTATGSCC 486
Db 421 CCGGAACCGAGATGCTGTACAGGTGACGCGGCTGTGCGCGCCAGTCCCTGTATGSCC 480
QY 487 CTGGAGTCCCGCGGCTGCGCAC 510
Db 481 CTGAAGTCCCGCGGCGCGCAC 504

RESULT 9
US-10-006-922-13
; Sequence 13, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Anemonia sulcata
US-10-006-922-13

Query Match 62.3%; Score 440.2; DB 14; Length 696;
Best Local Similarity 77.2%; Pred. No. 6.4e-96;
Matches 535; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 7 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCACTCGAGGGCACCGTGAACGGC 66
Db 4 GCTTCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTAAGGAGCGTTAATGGC 63

QY 67 CACTACTTCAAGTGCACCGGAGGCGAGGGCAACCCCTCGAGGGCAACCGAGGATG 126
Db 64 CACTACTTCAAGTGTACAGAAAGGAGAGGCAACCATTTGAGGGTACGCGAGAAATG 123

QY 127 AAGATCGAGGTGATCGAGGGGCGCCCTGCGCTTCGCCTTCACATCTCTCCACCTCC 186
Db 124 AAGATGAGGGTCAATCGAAGGAGGTCCATTTGCCATTTGCCCTTCACATTTTGTCAACAGT 193

QY 187 TGCATGTACGGCTCCAAGGCGCTTCATCAAGTACGTGTGTCGGCATCCCGACTACTTCAAG 246
Db 184 TGTATGTACGGTGTAGTAAAGGCTTCATCAAGATATGTGTACAGGAATTCCTGACTACTTCAAG 243

QY 247 CAGTCCCTCCCGAGGGCTTCACTGGAGCGCACCACTACGAGGACGGCGGCTTC 306
Db 244 CAGTCTTCCCTGAAAGTTTACTTTGGGAAAGAACCAACACTACGAGATGGAGGCTTT 303
QY 307 CTGACCGCCACACAGGACACTCCTCGACGGGACTGCTGTGTACAAAGGTCAAGATC 366
Db 304 CTTACAGTCTCATCAGGACACAGCCTAGATGAGATTGCCCTCGTTTCAAGGTCAAGATT 363
QY 367 CTGGCAACAACTTCCCGCGGACGCGCCCGTGTATGAGAAACAAAGGCGCGCGTGGAG 426
Db 364 CTTGGTAACAACTTCCCGCGGACGCGCCCGTGTATGAGAAACAAAGGTCGCGCGTGGAG 423
QY 427 CCTCCACCGAGATGCTGTACAGGTGACGCGGCTGTGCGCGCCAGTCCCTGTATGSCC 486
Db 424 CCATCCACCGAGATGTTTATGAAGTTGACGGTTCCTGCGGTGACAGTCTTTGATGGCC 483
QY 487 CTGGAGTCCCGCGGCTGCGCACCTGACCTGCCACCTGCCACACCACTACCGTCCCAAG 546
Db 484 CTTAAGTCCCTGGTGGTCTGCTCATCTGACTTGGCATCTCCATCTACTTACAGTCCAAA 543
QY 547 AAGCCCGCTCCGCTGAAAGATGCCCGCTTCCACTTCGAGGACCAACCGATCGAGATC 606
Db 544 AAACCAAGTGTGCTGCTTGAAGATGCCAGGATTTCAATTTGAAGATCATCGCATCGAGATA 603
QY 607 CTGGAGGAGGTGGAAGGCAAGGTCTACAAGCAGTACGAGGCGCGCTGGCGGCTAC 666
Db 604 ATGGAGGAAGTTGAGAAAGGCAAGTGTCTATAACAGTACGAGGACGAGTGGGCAAGTAC 663
QY 667 TCGAGCGCGCGCCCTCCCAAGCTGGGCGCAAC 699
Db 664 TGTGATGCTGCTCCATCCCAAGCTTGGACATAAC 696

RESULT 10
US-10-081-864-9
; Sequence 9, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Anemonia sulcata
US-10-081-864-9

Query Match 62.3%; Score 440.2; DB 15; Length 696;
Best Local Similarity 77.2%; Pred. No. 6.4e-96;
Matches 535; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 7 GCCTCCCTGTCGACCGAGACCATGCCCTTCAGGACCACTCGAGGGCACCGTGAACGGC 66
Db 4 GCTTCCTTTTAAAGAGACTATGCCCTTTAAGACGACCATTAAGGAGCGTTAATGGC 63

QY 67 CACTACTTCAAGTGCACCGGAGGCGAGGGCAACCCCTCGAGGGCAACCGAGGATG 126
Db 64 CACTACTTCAAGTGTACAGAAAGGAGAGGCAACCATTTGAGGGTACGCGAGAAATG 123

Qy	127	AAGATCGAGGTGATCGAGGGCGGCCCCCTGCGCTTGCCTTCCACATCTCTGTCCACCTCC	186
Db	124	AAGATAGAGGTTCATCGAAGGAGGTCCATTCGCAATTTGCCTTCCACATTTTGTCAACGAGT	183
Qy	187	TGATGTACGGCTCCAAAGSCCTTCATCAAGTACCTGTCCGSCATCCCCGACTACTTCAAG	246
Db	184	TGTATGTACGGTAGTAAGSCCTTCATCAAGTATGTGTCAAGAAVTCCTGACTACTTCAAG	243
Qy	247	CAGTCCCTCCCGAGGGGTTCACTGGGAGCGCACCACTACGAGGACGGCGGCTTC	306
Db	244	CAGTCTTCCCTGAAGTTTACTTGGGAAGAACCACTACGAGGATGGAGGCTTT	303
Qy	307	CTGACCGCCCAACAGACACCTCTCTGACGGGCACTGCCTGTGTAAAGGTGAAGATC	366
Db	304	CTTACAGCTCATCAGACACAAAGCTAGTAGGAGATTGCCTCGTTTACAAGGTCAAGATT	363
Qy	367	CTGGGCAACAACTTCCCGCGACGGCCCGTGTACGACAGGCGCGCCCTCGGAG	426
Db	364	CTTGGTAAATTTTTCCTGCTATGGCCCCGTGTATGAGAACAAAGCAGGAAGATGGAG	423
Qy	427	CCCTCCACCGAGATCGTGTACAGAGTGACGGCGTGTGGCGGCCAGTCCCTGATGGCC	486
Db	424	CCATCCACCGAGATAGTTTATGAAGTTGACGGTGTCTCGGTGGACAGCTCTTTGATGGCC	483
Qy	487	CTGAGATGCCCGCGGTGCGCACTGACCTGCCACCTGCACACCACTTACCGCTCAAG	546
Db	484	CTTAAGTGCCCTGTGTGTCATCTGACTTGGCACTCTCCATCTACTACTACAGTCCAAA	543
Qy	547	AAGCCGCGCTCGCCCTGAGAGTGCCTGCTTCCACTTCGAGGACACCCGATCTCGATC	606
Db	544	AAACCAAGCTGCTGCCTTGAAGATGCCAGATTTCAATTTGAAGATCATCGCATCTGAGATA	603
Qy	607	CTGGAGGAGGTGGAGAAGGGCAAGTGTCTAACAGCAGTACGAGGCGCGCTGGGCGCGTAC	666
Db	604	ATGAGGAAGTTGAGAAAGGCAAGTGTCTAAAAGTACGAGACGAGCAGTGGGCAGGTAC	663
Qy	667	TGCGACGCGCCGCCCTCCAAAGCTGGGCGCAAC	699
Db	664	TGTGATGTGTCTCCATCCAAAGCTTGGACATAC	696

```

RESULT 11
US-10-724-178-1052
; Sequence 1052, Application US/10724178
; Publication No. US20040137528A1
; GENERAL INFORMATION:
; APPLICANT: Odyssey Thera, Inc.
; APPLICANT: Michnick, Stephen
; APPLICANT: Macdonald, Marnie
; APPLICANT: Lamerdin, Jane
; TITLE OF INVENTION: FRAGMENTS OF FLUORESCENT PROTEINS FOR PROTEIN-FRAGMENT
; TITLE OF INVENTION: COMPLEMENTATION ASSAYS
; FILE REFERENCE: ODDY007
; CURRENT APPLICATION NUMBER: US/10/724,178
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: US 60/461,133
; PRIOR FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1052
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modif. frag.; KFP F1C, with position 1 Met removed
; NAME/KEY: CDS
; LOCATION: (1)..(456)
; OTHER INFORMATION: KFP F1C corresponds to aa residues 1-153 of KPPI
US-10-724-178-1052

```

Matches	445; Conservative	0; Mismatches	11; Indels	0; Gaps	0;
Qy	7	GCCTCCCTGCTGACCGAGACCATGCCCTTCAGACACACCATCGAGGGACCGTGAACGGC	66		
Db	1	GCCTCCCTGCTGACCGAGACCATGCCCTTCAGACACACCATCGAGGGACCGTGAACGGC	60		
Qy	67	CACCTACTTCAAGTGCACCGGCGAGGGCGAGGGCAACCCCTCGAGGGGCACCCAGGAGATG	126		
Db	61	CACCTGCTTCAAGTGCATCGGCAGAGGGCGAGGGCAACCCCTTCGAGGGGACCCAGGAGATG	122		
Qy	127	AAGATCGAGGTGATCGAGGGCGGCCCTGCCCTTCGCTTCCACATCTCTGTCCACCTCC	186		
Db	121	AAGATCGAGGTGATCGAGGGCGGCCCTGCCCTTCGCTTCCACATCTCTGTCCACCTCC	180		
Qy	187	TGCATGTACGGCTCCAAGGCTTCATCAAGTAGTGTCCGGCATCCCGGACTACTTCAAG	245		
Db	181	TGCATGTACGGCTCCAAGACCTTCATCAAGTAGTGTCCGGCATCCCGGACTACTTCAAG	240		
Qy	247	CAGTCCCTCCCGAGGGCTTTCACCTGGGAGCGCACCAACCACTACGAGGACGGCGGCTTC	306		
Db	241	CAGTCCCTTCCCGAGGGCTTTCACCTGGGAGCGCACCAACCACTACGAGGACGGCGGCTTC	300		
Qy	307	CTGACCGGCCACACAGGACACTTCCTGGACGGCGACTGCCTGTGTACAAAGTGAAGATC	366		
Db	301	CTGACCGGCCACACAGGACACTTCCTGGACGGCGACTGCCTGTGTGTACAAAGTGAAGATC	360		
Qy	367	CTGGGCAACAACATTTCCCGCCGACGGCCCCGTGATGCAGAACCAAGGCCGCGCTGGGAG	426		
Db	361	CTGGGTAAACAATTTCCCGCCGACGGCCCCGTGATGCAGAACCAAGTTCGCGCTGGGAG	420		
Qy	427	CCCTTCCACCGAGATCTGTGTACGAGGTGGACGGCGTG	462		
Db	421	CCCGGAAACCGAGATCTGTGTACGAGGTGGACGGCGTG	456		

```

RESULT 12
US-10-155-809-1
; Sequence 1, Application US/0155809
; Publication No. US2003092884A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Chudakov, Konstantin
; APPLICANT: Chudakov, Dmitry
; TITLE OF INVENTION: Kindling Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Their Use
; FILE REFERENCE: CLON-073
; CURRENT APPLICATION NUMBER: US/10/155,809
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/293,752
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/329,176
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/210,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/976,673
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 767
; TYPE: DNA
; ORGANISM: anthozoa
US-10-155-809-1

```

Query Match	61.8;	Score 436.8;	DB 15;	Length 767;
Best Local Similarity	76.7;	Pred. No. 4.2e-95;		
Matches 534;	Conservative 0;	Mismatches 162;	Indels 0;	Gaps 0;
Qy	7	GCCTCCCTGCTGACCGAGACCATGCCCTTCAGGACCACTCGAGGCGACCGTGAAACGGC	66	
Dy	72	GCCTCCCTTTTAAAGAAGACTATGCCCTTTAAAGACGACCATTCGAAGGAGCGGTTAATGGC	131	

; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/210,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/976,673
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 767
; TYPE: DNA
; ORGANISM: anthozoa
US-10-155-809-5

Query Match 51.3%; Score 433.6; DB 15; Length 767;
Best Local Similarity 76.4%; Pred. No. 2.4e-94;
Matches 532; Conservative 0; Mismatches 164; Indels 0; Gaps 0;
QY 7 GCCTCCTGCTGACCGAGACCATGCTCCCTTCAGGACCAACCATCGAGGACCGGTGAACGCG 66
Db GCTTCCTTTTAAAGAGACTATGCCCTTTAAGACCACTTGAAGGACGGTTAATGCG 131
QY 67 CACTACTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGCAACCCAGAGATG 126
Db 132 CACTACTTCAAGTGTACAGGAAAGGAGGGCAACCCATTTGAGGGTACGAGGAAATG 191
QY 127 AGATGAGGTGATCAGGAGGGGCGCCCTCGCCCTTCGCTTCCACATCTGCTCCACTCC 186
Db 192 AGATGAGGTGATCAGAGAGGTCCATTCGCCATTTGCCCTTCCACATTTTGTCAAGAT 251
QY 187 TGATGTAGCGCTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGACTACTTCAAG 246
Db 252 TGATGTAGCGTGTAAAGACCTTCATCAAGTATGTGTCAAGAAATTCCTGACTACTTCAAG 311
QY 247 CAGTCCCTCCCGAGGGCTCACTGGAGCGCAACCACTACAGGACCGCGCTTC 306
Db 312 CAGTCTTTCCTGAAAGTCTTACTTGGGAAAGAACCAACCACTTACAGGATGGAGCTTT 371
QY 307 CTGACCGCCCAACAGGACACTCCCTCGAGCGGCACTGCTGTGTACAAAGTGAAGATC 366
Db 372 CTTACAGCTCATCAGGACACAAGCTTAGATGGAGATTGCTCGTTTACAAAGTCAAGATT 431
QY 367 CTGGCAACAACCTTCCCGCGGAGCGGCGGCTGTATGAGAGAAAGCAAGCGCGCGCTGGAG 426
Db 432 CTTGGTAATAATTTCTGCTGATGCGCGGCTGTATGAGAGAAAGCAAGCAAGATGGAG 491
QY 427 CCCTCCACCGAGATCGTGTACGAGGTGAGCGGCTGTGCGGCGGCACTCCCTGTATGGC 486
Db 492 CCAGGCAACGAGATAGTTTGAAGTTGACGGTGTCTGCTGGGACAGTCTTTTATGGCC 551
QY 487 CTGGAGTGCCTCCCGGCTGCGCACTGACCTGACCTGACCACTGACCACTACCGCTCCAAG 546
Db 552 CTTAAGTGCCCTGGTGTGCTATCTGACTTGGCATCTCCATCTACTTACAGTCCAAA 611
QY 547 AAGCGGCTCCCGCTGAGATCCCGGCTTCCACTTCGAGGACCAACCGATCGAGATC 606
Db 612 AAACCAAGCTAGTGCCTTTGAAGTGGCAGGATTTTATTTGAAGATTACCGCATCGAGATA 671
QY 607 CTGGAGAGGTGGAGAGGGCAAGTGTCTCAAGCAGTACGAGCGCGCGCTGGCGCGCTAC 666
Db 672 ATGAGGAAGTTGAGAAAGGCAAGTGTCTAATACAGTACGAGCAGCAGTGGGCAAGTAC 731
QY 667 TGGAGCGCGCGCTCCCAAGCTGGGCAACACTGA 702
Db 732 TGTGATGCTCTCCATCCAAAGCTTGGACATACTAA 767

RESULT 15
US-09-976-673-11
; Sequence 11, Application US/09976673
; Patent No. US20020160473A1
; GENERAL INFORMATION:

; APPLICANT: Lukyanov, Sergey
; APPLICANT: Pradkov, Arcady
; APPLICANT: Labas, Yulii
; APPLICANT: Matz, Mikhail
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Gurskaya, Nadezda
; TITLE OF INVENTION: FAR RED SHIFTED FLUORESCENT PROTEINS
; FILE REFERENCE: CLON-028WO
; CURRENT APPLICATION NUMBER: US/09/976,673
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,018
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: 60,306,131
; PRIOR FILING DATE: 2001-07-16
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 687
; TYPE: DNA
; ORGANISM: heteractis crispa
US-09-976-673-11

Query Match 59.1%; Score 417.8; DB 9; Length 687;
Best Local Similarity 79.1%; Pred. No. 1.5e-90;
Matches 523; Conservative 0; Mismatches 132; Indels 6; Gaps 2;
QY 12 CCTGTGACCGAGACCATGCTCCCTTCAGGACCAACCATCGAGGGCACCGTGAACGCCACTA 71
Db 12 CCTGCTGAAGAGAGTATGCGCATCAAGTATACATGAGGGGACCGTGAACGCCACTA 71
QY 72 CTTCAAGTGCACCGGCAAGGGGAGGGCAACCCCTCGAGGGGACCCAGGAGATGAAGAT 131
Db 72 CTTCAAGTGCAGGGGCGAGGGGCGGCAACCCCTTCGCGGGACCCAGGAGATGAAGAT 131
QY 132 CAGGTGATCAGGCGGCGCCCTCGCTTCCATCTCCATCTCTGCTCCACTCTCTGAT 191
Db 132 CCAGTGCACCGGCGGCGCCCTCGCTTCCATCTCTGCTTCCATCTCTGCTGCGA 191
QY 192 GTACGGTCCAAAGGCTTCATCAAGTACGTGTCCGGCATCCCGCACTACTTCAAGCAGTC 251
Db 192 GTACGGCAGCAGGACCTTCGTGCAACCAACCGCGGAGATCCCGCACTTCTTCAAGCAG 251
QY 252 CTTCCCGGAGGCTTCACCTGGGAGAGAACCAACCACTTACGAGGACGGCGCATCTTGAC 311
Db 252 CTTCCCGGAGGCTTCACCTGGGAGAGAACCAACCACTTACGAGGACGGCGCATCTTGAC 311
QY 312 CGCCCAACAGGACACTCCCTGGACGGCACTGCTGTGTACAAAGTGAAGATCTCTGGG 371
Db 312 CGCCCAACAGGACACTCCCTGGAGGGCACTGCTGTGTACAAAGTGAAGTGCACGG 371
QY 372 CAACACTTCCCGCGGAGCGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 431
Db 372 CACCAACTTCCCGCGGAGCGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 431
QY 432 CACCGAGATCGTGTACGAGGTGGACGGCTGCTGCGCGGCGAGTCCCTGTATGGGCTTGA 491
Db 432 CACCGAGGTGTGTACCCCGGAGACGGGTGCTGTGCGCGGAGAACGCTGTATGGGCTTGA 491
QY 492 GTGCCCCGGCGGTGCGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 551
Db 492 G---GTGGCGGACCGGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 548
QY 552 GCGCTCCCGCTGAGATGCGCGGCTTCCACTTTCGAGGACCAACCGCATCGAGATCTCTGA 611
Db 549 CTTGCGCGCTTACCATGCGCGGCTTCCACTTTCACCGATGATGCGGCTCCAGATGCTGCG 608
QY 612 GGAGGTGGAGAGGGCAAGTGTCTAAGCAGTACGAGGCGCGCGCTGGGCGCGCTACTGCGA 671
Db 609 GAA---GAAGAGGACGAGTACTTTCGAGCTGTACGAGGCGGCGGCTGGGCGCGCTACAGCA 665
QY 672 C 672
Db 666 C 666

Wed Aug 4 09:01:05 2004

us-10-081-864-21.rnpb

Page 10

Search completed: August 3, 2004, 19:01:43
Job time : 516.043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2004, 11:02:24 ; Search time 3167.97 Seconds
(without alignments)
6664.376 Million cell updates/sec

Title: US-10-081-864-21

Perfect score: 707
Sequence: 1 ggcacgcctccctgctgac.....ctgggccaactgaagctt 707

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513299 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estov:*
5: em_estpl:*
6: em_estro:*
7: em_estro:*
8: gb_est1:*
9: gb_est1:*
10: gb_est2:*
11: gb_est2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101.2	14.3	550	13 BQ551001	BQ551001 H4006G01-
2	73.6	10.4	562	14 CA937115	CA937115 sav29e05.
3	72.2	10.2	625	12 BG906607	BG906607 TaLr1151A
4	71.8	10.2	749	14 CF489659	CF489659 POL1_58_H

5	71	10.0	852	14	CD438780	CD438780	EL01N0517
6	59.8	9.9	549	12	BI943903	BI943903	BI943903 ea52309.Y
7	59.4	9.8	606	12	BI942400	BI942400	BI942400 sb80112.Y
8	59.4	9.8	986	12	BM561582	BM561582	BM561582 AGENCOURT
9	59.4	9.8	990	10	BE216934	BE216934	BE216934 EST0477.T
10	69	9.8	593	29	CG332384	CG332384	CG332384 OGVDD58TV
11	68.8	9.7	697	13	BQ578547	BQ578547	BQ578547 UI-H-FTO-
12	67.6	9.6	496	13	BQ578547	BQ578547	BQ578547 WHE0306.E
13	67.6	9.6	631	14	CA754991	CA754991	CA754991 BR0300100
14	67.6	9.6	695	14	CD902570	CD902570	CD902570 G356.107G
15	67.4	9.5	869	14	CK159167	CK159167	CK159167 FGAS04056
16	67.2	9.5	925	29	CNS0001P	CNS0001P	AL033013 Drosophila
17	66.6	9.4	545	12	BI974926	BI974926	BI974926 sal74g12.
18	66.2	9.4	799	14	CB656196	CB656196	CB656196 OSJNEC10E
19	66	9.3	753	29	CG231715	CG231715	CG231715 OGVG062TH
20	66	9.3	835	29	CG51863	CG51863	CG51863 OGVBR82TC
21	66	9.3	842	29	CG273302	CG273302	CG273302 OGVBO38TV
22	66	9.3	2299	11	AY106831	AY106831	AY106831 Zea mays
23	65.4	9.3	715	10	BE519855	BE519855	BE519855 HV_CBE002
24	65.4	9.3	728	14	CF484651	CF484651	CF484651 POL1_26.D
25	65.2	9.2	982	13	BX415111	BX415111	BX415111 BX415111
26	65	9.2	473	14	CD86846	CD86846	CD86846 G118.103H
27	64.8	9.2	672	14	CA501436	CA501436	CA501436 WHE4033.F
28	64.8	9.2	674	13	BQ255271	BQ255271	BQ255271 WHE2868-B
29	64.8	9.2	822	29	CG301217	CG301217	CG301217 OGVFR69TH
30	64.8	9.2	852	29	CG311347	CG311347	CG311347 OGVBC17TC
31	64.8	9.2	929	29	CC719024	CC719024	CC719024 OGGAC14TV
32	64.4	9.1	624	6	AL811149	AL811149	AL811149 Triticum
33	64.2	9.1	538	12	BM178969	BM178969	BM178969 saj61f08.
34	64.2	9.1	644	14	CF486437	CF486437	CF486437 POL1_37.E
35	64.2	9.1	647	14	CD233835	CD233835	CD233835 S81_4.F02
36	63.8	9.0	967	13	BX428434	BX428434	BX428434 BX428434
37	63.8	9.0	1079	14	CK210801	CK210801	CK210801 FGAS02262
38	63.6	9.0	546	12	BI780772	BI780772	BI780772 Ebes01.SQ
39	63.6	9.0	622	12	BG839331	BG839331	BG839331 Gm01_14e0
40	63.6	9.0	778	14	CF480713	CF480713	CF480713 POL1_67.A
41	63.6	9.0	792	14	CF488222	CF488222	CF488222 POL1_48-B
42	63.6	9.0	1147	14	CK209786	CK209786	CK209786 FGAS02156
43	63.4	9.0	822	29	CG337308	CG337308	CG337308 OGVFR58TV
44	63.4	9.0	877	29	CG337294	CG337294	CG337294 OGVFR58TV
45	63.4	9.0	885	29	CC727899	CC727899	CC727899 OGKAO61TH

ALIGNMENTS

RESULT 1
BQ551001
LOCUS H4006G01-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
DEFINITION 550 bp mRNA linear EST 20-JUN-2002
ACCESSION BQ551001
VERSION BQ551001
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 550)
AUTHORS VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G.,
Martin,P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T.,
Kargul,G.J., Luo,A.G., Kelso,J., Hide,W. and Ko,M.S.H.
TITLE Assembly, verification, and initial annotation of NIA 7.4K mouse
JOURNAL CDNA clone set
MEDLINE Genome Res. 12 (12), 1999-2003 (2002)
PUBMED 22354164
COMMENT Other ESTs: H4006G01-3
Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please visit http://lgsun.grc.nia.nih.gov/cDNA/NIA_7_4k.html for details.
 Plate: H4006 row: G column: 01
 Seq primer: -21M13 Reverse
 High quality sequence stop: 550
 POLYA=No.

FEATURES

Location/Qualifiers

source

1. 550 /organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="niaEST:H4006G01-5"

/db_xref="taxon:10090"

/clone="H4006G01"

/sex="mixed"

/dev_stage="mixed"

/lab_host="DH10B"

/note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; This

clone is among a rearranged set of 7,407 clones from more than 20 cDNA libraries."

ORIGIN

Query Match 14.3%; Score 101.2; DB 13; Length 550;
 Best Local Similarity 57.4%; Pred. No. 0.00023;
 Matches 225; Conservative 0; Mismatches 158; Indels 9; Gaps 2;
 QY 13 CTGCTGACCGGAGACCATGCGCTTCAGGACACCACTGAGGCGACCGGTGAACGGCCACTAC 72
 DB 159 CTGTTACCGGGGTGGTCCCATCTCTGTCAGCTGACGCGGACGTAAACGGCCACAG 218
 QY 73 TTCAAGTGCACCGCAAGGGGAGGCAACCCCTCGAGGGACCCAGGAGTGAAGATC 132
 DB 219 TTCAAGCTGTCCGCGAGGGGAGGCGATGCCACCTACGCGAAGCTGACCTGAAGTTC 278
 QY 133 GAGGTGATCGAGGCGCGCCCTCGCTTCGCTTCCACATCTGTCACCTCTGTCATG 192
 DB 279 ATCTGACACAC---CGGAGCTGCTCCGCTGCGCCACCTCTGTCACCTCTGTCATG 335
 QY 193 TACGGCTTCAAGGCTTTCATCAAGTACGTGTCGGCAT-----CCCGACTACTTCAAG 246
 DB 336 TACGGGTGACGTCTTTCAGCGGTACCCGACACGTGAAGCAGCAGCACTTCTTCAAG 395
 QY 247 CAGTCCCTCCCGAGGCTTCACTCGGAGGACACCACTACGAGGACGCGGCTTC 306
 DB 396 TCCGCCATGCGGAGGCTAGCTCAGGAGCGCACCATCTTCTCAAGGACGACGCAAC 455
 QY 307 CTGACCGCCGCCACGAGCACCTCCCTCGAGCGGACCTGCTGCTGTACAGGTGAAGATC 366
 DB 456 TACAGACCCGCGCGGAGGTGAAGTTCGAGGGGACACCTCTGTGAACCCGATCGAGCTG 515
 QY 367 CTGGGGAACAACTTCCCGCGGACGCGCCCGT 398
 DB 516 AAGGGCATCGACTTCAAGGAGGAGCGGCAACAT 547

RESULT 2

CA937115

LOCUS

DEFINITION CA937115 562 bp mRNA linear EST 30-DEC-2002
 sav29e05.y1 Gm-c1048 Glycine max cDNA clone SOYBEAN CLONE ID:
 Gm-c1048-6754 5' similar to SW:ER1_HEVER Q39963 ETHYLENE-INDUCIBLE
 PROTEIN HEVER. ; mRNA sequence.

CA937115

VERSION

CA937115.1

GI:27425595

KEYWORDS

EST.

SOURCE

Glycine max (soybean)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Glycine.

REFERENCE

1 (bases 1 to 562)

AUTHORS

Shoemaker, R., Reim, P., Vodkin, L., Erpelding, J., Coryell, V.,

Khanna, A., Bolla, B., Marra, M., Hallier, L., Kucaba, T., Martin, J.,
 Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
 Bowers, J., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
 Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
 McCann, R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact: ccu@resgen.com web site:
 www.resgen.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 407.

FEATURES

Location/Qualifiers

source

1. 562 /organism="Glycine max"

/mol_type="mRNA"

/db_xref="taxon:3947"

/clone="SOYBEAN CLONE ID: Gm-c1048-6754"

/tissue_type="whole seedlings of greenhouse grown plants"

/dev_stage="1 week old"

/lab_host="DH10B"

/clone_lib="Gm-c1048"

/note="vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
 XhoI; The Clark NIL was constructed and seed was provided
 by Dr. J. Specht, University of Nebraska (Shoemaker and
 Specht, 1995). The cDNA library was constructed from mRNA
 isolated from whole seedlings of 1 week old greenhouse
 grown plants. Complementary DNA was synthesized from mRNA
 using a primer consisting of a poly(dT) sequence with a
 XhoI restriction site and a 3' anchor. EcoRI adapters were
 ligated to the blunt-ended cDNA fragments followed by XhoI
 digestion. The cDNA fragments were directionally cloned
 into the EcoRI-XhoI restriction site of the pBluescript
 vector. The ligated cDNA fragments were transformed into
 DH10B host cells (GibcoBRL). The library was constructed
 in cooperation with Dr. Paul Keim's laboratory at Northern
 Arizona University."

ORIGIN

Query Match 10.4%; Score 73.6; DB 14; Length 562;
 Best Local Similarity 47.0%; Pred. No. 1.2;
 Matches 263; Conservative 0; Mismatches 294; Indels 3; Gaps 1;
 QY 36 CAGGACCAACCATCGAGGACACCGTGAACGGCCACTACTTCAAGTGCACCGGCAAGGGCGGA 95
 DB 1 CACGATCCCGCTGATGGCCAAAGGCCCGCATCGGCCACTTCTGTCGAAGCCGACATCCTCGA 60
 QY 96 GGGCAACCCCTCGAGGGGACCCAGGAGATGAAGATCGAGTGTAGCGGGCGGCCCT 155
 DB 61 AGCATCGGCATCGACTACTGTGACGAGAGGAGGTCTCTACCTCCCGACGACGGCAA 120
 QY 156 GCCCTTCGCTTCCACATCTCTGCCACTCTCTGTGATGTACGGTCTCAAGGCTTCATCAA 215
 DB 121 CCACATCAACAAACACAACTTCCGCATCCCTTCTGCTGTGGTCCGCAACCTCGCGCA 180
 QY 216 GTACGTCTCGGCATCCCGGACTTCAAGCATCTCTCCCGAGGCTTCACCTGGGA 275
 DB 181 GGCCCTCCGCCCATCCCGAGGGCGCGCATATCCGACCAAGGGCGAGGCGCGCAC 240
 QY 276 CGGCACCAACCACTACGAGGACGCGGCTTCTTCAACCGCCACCCAGGACACCTCCCTGGA 335
 DB 241 CGGCAACATCATCGAGCGCTTCCGACCGCTCCGCTCGTCTATGAGCGACATTAGGGTCT 300
 QY 336 CGGCGACTGCTGTGTACAGGTGAAGATCTCTGGGCAACACTTCCCG---CCGACGG 392

Db 301 CCGCAACATGGACGACGAGGCTTTCACCTTCGCCAAGAACATCGCTCGCCCTATGA 360
 QY 393 CCCCGTATGACAGAACAGCCCGCGCTGGAGCCCTCCACCGAGATCGTATACAGGT 452
 Db 361 CTTCTGTCATGACAGCAAGAGCTTGGCGCGCTCCCGCTTGTCCACTTTCGCGCGCGCG 420
 QY 453 GGACGGGTCTCGCGCGCCAGTCCCTGATGGCCCTGGAGTCCCGCGCGCTCGCCACCT 512
 Db 421 AGTGGCACCCCGCGCGACCGCGCTCATGATGACGCTCGGTGCGAGCGGTCTTCTGT 480
 QY 513 GACCTGCGACCTGCACACACCTACCGCTCCAAAGAGCCCGCTCCCGCCCTGAAGATGCC 572
 Db 481 CGGCTCGCGGTCTTCAAGAGCGCGATCCCGGAGCGCGCCAGGCGGATTTGTCAGGC 540
 QY 573 CGGCTTCCACTTCAGGACC 592
 Db 541 CGTGACTCACTACAGCGACC 560

RESULT 3

BG906607 625 bp mRNA linear EST 05-JUN-2001
 LOCUS Talr1151A09R Talr1 Triticum aestivum cDNA clone Talr1151A09 5',
 DEFINITION mRNA sequence.

ACCESSION BG906607

VERSION BG906607.1 GI:14314283

KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)

ORGANISM

REFERENCE

AUTHORS Cloutier, S., Dong, G. and Walsh, A.

TITLE Wheat functional genomics - Thatcher Lr1 cDNA library

JOURNAL Unpublished (2001)

COMMENT Contact: Dr. Sylvie Cloutier

Cereal Research Centre, Agriculture and Agri-food Canada

195 Dafee Rd, Winnipeg, MB, Canada R3T 2M9

Tel: (204) 983-2340

Fax: (204) 983-4604

Email: scloutier@agr.gc.ca

was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end). Average insert size is >2.2 kb

Plate: 151 row: A column: 09

Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers

1..625

/organism="Triticum aestivum"

/mol_type="mRNA"

/cultivar="Thatcher Lr1"

/db_xref="taxon:4565"

/clone="Talr1151A09"

/tissue_type="Leaf tissue"

/dev_stage="14 Days old"

/lab_host="E. coli XL0UR"

/clone_lib="Talr1"

/notes="Vector: Lambda ZapII; mass excised in plasmid vector pBK-CMV (Stratagene).; Site 1: EcoRI; Site 2: XhoI; mRNA obtained from wheat NIL Thatcher Lr1 24 hours after inoculation with leaf rust pathogen Puccinia triticiina race BBB carrying the avirulence gene Avr1."

ORIGIN

Query Match 10.2%; Score 72.2; DB 12; Length 625;
 Best Local Similarity 47.7%; Pred. No. 1.8;
 Matches 243; Conservative 0; Mismatches 263; Indels 3; Gaps 1;
 QY 87 CAGGGCGAGGGCAACCCCTCGAGGCGACCCAGGAGTGAAGATCGAGTGATCGAGGG 146
 Db 11 CGAGCTCGAGTCCATCGGGGGGACTACGTGGAGAGAGGAGGTCTCACGCTCGCCGA 70

FEATURES

source

Seq primer: Sug5 (CTTCTGCTCTAAAGCTGGG).

Location/Qualifiers

1..749

/organism="Sorghum bicolor"

QY 147 CGGCCCCCTGCGCTTCACATCTCTGCACCTCTGATGTACGGCTCAAGGC 206
 Db 71 CGAGCCCAACACATCAACAAGCAACTTCGGGTGCCCTTCGTCTCGGCTGACGAA 130
 QY 207 CTTTCATCAAGTACGTTCGGCATCCCGACTACTTCAAGCAGTCCCTTCCCGAGGGCTT 266
 Db 131 CTTGGCGAGGCCCTTCGCGCATCCGCGAGGGCGGCCATGATCCGCAACCAAGGCGA 190
 QY 267 CACTGGGAGGCGACCACTAGAGGAGCGGGCTTCTGACCGCCACACAGGACAC 326
 Db 191 GCGCGCACCGGCAATGTATCGAGGCGCTCCGCAACGTCGCTCCGTCATGGGCGAGCT 250
 QY 327 CTCCTGAGACGGCGACTGCTGCTGTACAAGTGAAGATCTGGGCAACAACATTTCCCCGC 386
 Db 251 CGGTGCCCTTCGAGCATGACGACGAGGATTTACCTATGCCAAGACATCGCGCG 310
 QY 387 CG--ACGGCCCCGTGATGACAGAACAGGCGCGCGCTGGAGCCCTCCACCGAGATCGT 443
 Db 311 ACCGTACGACCTCGTTCATGACAGCAAGCAGCTCGGCGCTGCGCGCTGCTCCAGTTGCG 370
 QY 444 GTACAGGTGACGGCGTGTGCGGCGCACTCTGATGCGCTGGAGTCCCGCGGG 503
 Db 371 GCGCGCGGTGTGCGCACGCGCGCGATGCGGCGCTTATGATGAGCTCGGCTGGACGG 430
 QY 504 TCGCCACCTGACCTGCCACCTGCACACACCTACCGTCTCCAGAGAGCGCGCTCGGCCCT 563
 Db 431 TGTCTTGTGCTCGCTCGGTGTCTTCAAGAGCGGCGACCCCGCGCGCGCGCGCCAT 490
 QY 564 GAAGATCGCGCTTCCACTTCGAGGACC 592
 Db 491 CGTGAGGCCGTCCACCCACTACAGCGACC 519

RESULT 4

CF489659

LOCUS

DEFINITION

POU1_58_H06.g1_A002 Pollen Sorghum bicolor cDNA clone

CF489659

ACCESSION

VERSION

CF489659.1 GI:34518528

KEYWORDS

SOURCE

ORGANISM

Sorghum bicolor (sorghum)

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 749)

REFERENCE

AUTHORS

Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C.,

Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J.,

Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J., and

Pratt, D.H.

EST database from Sorghum: pollen

Unpublished (2003)

Other ESTs: POL1_58_H06.b1_A002

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in

the Human Genome Center, University of Tokyo Institute of Medical

Science; plant material and RNA prepared at Texas A & M University;

sequencing done in the Laboratory for Genomics and Bioinformatics,

University of Georgia. Sequence ends have been trimmed to exclude

vector and regions below Phred quality 16. Three-prime sequences

are presented as their reverse complement and have been trimmed to

exclude polyA.

Seq primer: Sug5 (CTTCTGCTCTAAAGCTGGG).

Location/Qualifiers

1..749

/organism="Sorghum bicolor"

clade: Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 852)
Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
Messing,J.
Sequencing of the maize endosperm ESTs
Unpublished (2002)
Contact: Lai, Jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T3.
Location/Qualifiers
1..852
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W22"
/db_xref="taxon:4577"
/tissue_type="Endosperm of 7-23DAP"
/clone_lib="Endosperm 5"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

ORIGIN

Query Match 10.0%; Score 71; DB 14; Length 852;
Best Local Similarity 47.5%; Pred. No. 2.8;
Matches 277; Conservative 0; Mismatches 300; Indels 6; Gaps 2;

QY 2 GATCGCCCTCCCTGTGACCGAGACCATGCCCTTCAGAGACACCATCGAGGACCGGTGA 61
DB 259 GGTGGGACTCCACCCCGCGCTGTCCAGGCTCGGCGGGCTCTTCGCGTCCACCGCC 318
QY 62 ACGGCGACTACTTCAAGTGCACCGGCAAGGGCGAGGCAACCCCTTCAGAGGACACCGAGG 121
DB 319 GCTGGCGGACGCCCTCGCGGTGTGAGCGCTACGGCTTCGGCCCGTCCACACCTCT 378
QY 122 AGATGAAGATGAGGTGATCGAGGCGCGCCCTGCTCCCTTCGCTTCCACATCTCTGTCCA 181
DB 379 CCATCTCTCTCCCGCGCACCCCGCGCTCTCTCTCTCGCGCGAGCACCGCGTGTCTCC 438
QY 182 CCTCTGTGATGTACGGCTCCAG--GCCTTCATCAAGTACGTGTCTCGGCTATCCCGACT 238
DB 439 CGCCCTGGGTGACCCCTCCGGCTCCGGCTCCGGCGCGGAGGAGCTCGGACCCCGACA 498
QY 239 ACTTCAAGCAGTCCCTCCCGAGGCTTTCACCTGGAGCGGACCAACCACTTACGAGGAG 298
DB 499 CCACCGCGCAGGTCCACCCCGGCTCTTTCACCAAGGCGCTCTCGCGCGTCCGAGCGG 558
QY 299 GCGGCTTCTGACCGCCACACAGACACCTCCCTGGAGCGGAGCTCCCTGTGTGTACAGG 358
DB 559 AGGTCTGATTCGCGGAGTGTGAGCGGTGTCTGTGGCGCGGAGCGGCGCTCGCGCGG 618
QY 359 TGAAGATCTGGGCAACAACTTCCCGCGCGAGCGGCTGATGAGAACAAAGGCGCGCC 418
DB 619 TCGTGTCAAGGGCGCGACCGCGTGTGAGCGCGGAGCGCGCTGTCTGTCTCGCTCGGCC 678
QY 419 GCTGGAGCCCTCCACCGAGATCTGTGTACGAGGTGAGCGGCTGTGTGCGCGGAGTCCC 478
DB 679 CGTGGTCCCGCGCTCCAGGTGTGTGAGCGGTGTGTGAGCGGTGTGTGAGCGGTCAAGGCG 738
QY 479 TGATGCCCTGGAGTGGCGCGCGGTTCGCTACCTGACCTGCACCTGCACACCACTTACC 538
DB 739 ACAGCATCTGTCTCGCGCGCGCGGAGCGCGA---GAAGGTACGCGGACTGCTCTTCC 795
QY 539 GTTCAAGAAGCCCGCTTCGCGCTTGAAGATGCGCGGTTCGA 581
DB 796 TCAGCTACCGCGGAGCGCGCGCGGCGGCAAGATGCTCGACCCGA 838

RESULT 6
B1943903

/mol_type="mRNA"
/cultivar="Brix623"
/db_xref="taxon:4558"
/clone="POL1.58.H06.A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Pollen"
/note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI;
Site 2: XhoI; The library was prepared from polyA+ RNA
from pollen at the late vacuolated-vacuolated stage of
development. Pollen was harvested from greenhouse-grown
panicles of sorghum line Brix623. Panicles were removed
from the flag leaf prior to emergence, when no detectable
amylase is present in pollen of male-fertile lines. This
stage represents pollen collected from anthers about 8-14
days prior to anthesis. Double-stranded cDNA was cloned
unidirectionally into different DraIII sites of the
pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG,
3-prime DraIII site is CACCATGTG). XhoI excises the cDNA
insert."

ORIGIN

Query Match 10.2%; Score 71.8; DB 14; Length 749;
Best Local Similarity 48.4%; Pred. No. 2.1;
Matches 233; Conservative 0; Mismatches 242; Indels 6; Gaps 1;

QY 9 CTCCTGTCTGACCGAGACCATGCCCTTCAGAGACACCATCGAGGCAACCGTGAACGCCA 68
DB 207 CTCGCTCGCGCGGTGCGCTCGGCTCCCGAGTACTCCGACTCCGCGCAAGGTGGCTA 266
QY 69 CTACTTCAAGTGCACCGGCAAGGGCGAGGGCAACCCCTCGAGGGGACCCAGAGATGAA 128
DB 267 CGTCTCCAGGCACTGCGACCTCGGGATCGTCTTCCGAGGCGGACCAAGGAGAGGT 326
QY 129 GATGAGGTGATCGAGGCGCGCCCTCGCTTCCGCTTCCACATCTGTCTCCACTCTCTG 188
DB 327 GGTGTCTGAGGAGCGGCGGTGGCTTCTCCCTTGGCGTCTGTGATGTTGGCA 386
QY 189 CATGTACGGTCTCAAGGCTTTCATCAAGTACGTGTCTCGGCAATCCCGACTACTTCAAGCA 248
DB 387 CAAGCCCGCCACCGGTGACCGGACCTGACGCTCTCTTCTCGCGGACACCTCCAAAGG 446
QY 249 GTCCCTCCCGAGGCTTCACTGGAGGCGGACCAACCACTTACGAGGACGCGGCTTCT 308
DB 447 CCACAAGCAGCGGCTTCACTTCCAGCCGCTCTGGCATCTGTAAGACCTCCAT 506
QY 309 GACCGCCACACGAGACACCTCCCTGGAGCGGACTGCTGTGTACAA-----GTTGAA 362
DB 507 CGCGTCCGACGGCTCGCCCAAGGACCGCGAGGCGATGCGCTCAACTGCTTGAGGC 566
QY 363 GATCTGGGCAACAACTTCCCGCGGAGGCGCGCTGTATGAGNAACAAGGCGGCGCGCTG 422
DB 567 GCGCTGAGCATGAGCATATCCCGCGGAGGCGCGGTGTGTCTCAACACCGGTGAACCT 626
QY 423 GGAGCCCTCCACCGAGATCTGTGTACGAGGTGAGCGGCTGTCTGCGCGGCACTCCCTGAT 482
DB 627 GCCCTTGTCAAGACGCTCGGCTCGGCGCGGACTCTGTCCGATCGACGCGACTCCAT 686
QY 483 G 483
DB 687 G 687

RESULT 5
CD438780
LOCUS
DEFINITION
E010N0517C04.b Endosperm_5 Zea mays cDNA, mRNA sequence.
ACCESSION
CD438780
VERSION
CD438780.1 GI:31354423
KEYWORDS
EST.
SOURCE
Zea mays
Zea mays
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com

Seq primer: -40RP from Gibco
High quality sequence stop: 428.

FEATURES

Location/Qualifiers
1..606

/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-ci016-6528"
/tissue_type="immature flowers of field grown plants"
/lab host="XL10-Gold"
/clone.lib="Gm-ci016"
/note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from immature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."

ORIGIN

Query Match 9.8%; Score 69.4; DB 12; Length 606;
Best Local Similarity 47.1%; Pred. No. 4.4;
Matches 248; Conservative 0; Mismatches 276; Indels 3; Gaps 1;
QY 69 CTACTTCAAGTGCACCGGAGGCGGCAACCCCTCGAGGCAACCCAGGAGTGA 128
Db 24 CCACTTCGTGAGCCAGATCTCTGAGCCATCGGATGATGATCGTGCAGGAGCGA 83
QY 129 GATCAGGTGATCAGGGGCGCCCTCGCCCTTCGCTTCCATGCTGTCACCTCTG 188
Db 84 GGTCTCACCTCCGCGAGCGCAACCATCAACAAACAACTTCGCGATCCCTT 143
QY 189 CATGTACGGTCCCAAGGCTTCATCAAGTACGTGTCGGCATCCCGACTACTCAAGCA 248
Db 144 CGTCTGTGCTGCGCAACCTCGCGAGGCGCTTCGCGCATCCGAGGGCGCGCCAT 203
QY 249 GTCCCTCCCGAGGGCTTCACCTGGAGCGCACCCACCTACGAGACGGCGGCTTCCT 308
Db 204 GATCCGCAACCAAGGCGGCGCACCGGCAACATCATCGAGGCGCTTCGCCACGTCG 263
QY 309 GACCGCCCAACGAGACCTCCCTGGAGCGGAGTGGCTGTGTACAGTGAAGATCCT 368
Db 264 CTCGCTCATGAGCGACATTAGGGTTCTCCGCAACATGAGACGAGGAGCTTTACCTT 323
QY 369 GGGCAACAATCTCCCGG---CCGAGCGCCCGTGTGATGACGAAACAGCGCGCTGGGA 425
Db 324 CGCCAGAACATGCTGCGCTTATGACCTGTCATGACGACCAAGCAGCTTGGCGCT 383
QY 426 GCGCTCCACGAGATCGTGTACAGGTGAGCGGGTGTCTCGCGGCGAGTCCCTGTATGGC 485
Db 384 CCGCTGTGTCCATCTCGCCCGCGGAGTGGCCACCCCGCGAGCGCGCTCATGAT 443
QY 486 CTTGAGTGTCCCGGCGGTGCGCACCTGACCTGCACTGCAACACCTACCGTCCAA 545
Db 444 GCAGTCCGTGCGAGCGGCTCTTCGCTCGGCTCGGCGTTTCAAGAGGGCGATCCCG 503
QY 546 GAAGCCCGCTTCGCCCTGAAGATGCCCGCTTCCACTTCGAGGACC 592
Db 504 GAAGCGCGCGAGGCGGATTGTCAGGCGGTGACTCACTACAGCGACC 550

RESULT 8
BM561582

LOCUS

DEFINITION

BM561582 986 bp mRNA linear EST 20-FEB-2002
AGENCOURT_6567315 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5739763
5', mRNA sequence.

ACCESSION

BM561582

VERSION

BM561582.1 GI:18806953

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 986)

AUTHORS

NIH-MGC http://mgc.mci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://imgc.llnl.gov

Plate: LLAM12753 row: j column: 20

High quality sequence stop: 680.

FEATURES

source

1..986

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5739763"

/tissue_type="duodenal adenocarcinoma, cell line"

/lab host="DH10B (phage-resistant)"

/clone.lib="NIH MGC 88"

/note="Organ: small intestine; Vector: pCMV-SPORT6;

Site 1: NotI; Site 2: SalI; Cloned unidirectionally;

oligo-dt primed. Average insert size 1.767 Kb. Library

enriched for full-length clones and constructed by Life

Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 9.8%; Score 69.4; DB 12; Length 986;
Best Local Similarity 45.1%; Pred. No. 4.7;
Matches 314; Conservative 0; Mismatches 371; Indels 12; Gaps 1;
QY 11 CCCTGCTCAGCAGACCATGCCCTTCAGGACCAACATCGAGGGCACCGTGAACGGCCACT 70
Db 64 CCAAGGTCGTGGAAACCAACCGGTGTGGGACAGGAACGGGGGTGCAGCCACCTGTGCT 123
QY 71 ACTTCAGTGCACGGGACGAGGGGAGGCAACCCCTCGAGGGCACCCAGGAGATGAAGA 130
Db 124 TCTTCACACCCACGCAACCCGGTGTGGCTGCCCATCGGCTTGAGTGACA 183
QY 131 TCGAGGTGATCGAGGG 190
Db 184 TGAAGACCTGATCGTGCCTGAGGCTTCCTGGCTTCACAGGAGCGGCTCCCTGCA 243
QY 191 TGTACGGCTCAAGGCTTCATCAAGTACGTGTGCGGCATCCCGACTACTTCAAGCAGT 250
Db 244 GGATCTCCCTCGAGACCAATAACACGACGCTGGGCTCCGCTCAGGGGGTCAAGGAG 303
QY 251 CCCTCCCGAGGGTTCACCTGGGAGGCGCACACCACTACGAGAGCGGGGCTTCCTGA 310
Db 304 CTTAGCCCTGGAATTGATGTGTCCAAACACACATCTACTGACAGAGCTCAGCTTGA 363
QY 311 CGGCCCAACAGGACCTTCCTCGTGAACGGGAGCTGCTGGTGTACAAAGTGAAGATCTGG 370
Db 364 AGACCATCAGCGGCGCTTCATGAACGGGAGCTCGGTGAGGACGCTGGTGGAGTTGGCC 423
QY 371 GCAACAACCTCCCGCGAGCGGCGGCTGATGAGAACAGGCGGCGGCTGGGAGCCCT 430
Db 424 TTGACTACCCGAGGAGTATGCGCTTACTGATGGCAAGAACCTTACTTGGGCGGACA 483

```

QY 431 CCACCGAGATGCTGTAGACGGTGAACGGCGTG-----CTGCGCGGCGCAGTCCC 478
Db 484 CTGGGACCAACAGAAATCGAGTGGCGGCTGGACGGCGAGTTCCGGCAAGTCTCGTGT 543
QY 479 TGATGGCCCTGGAGTGCCTGGCGGCTGCGCACCTGACCTGCCACCTGSCACACACCTACC 538
Db 544 GGAGGACTTGGACAAACCGAGGTGCTGSCCTTGGATCCACCAAGGGTACATCTACT 603
QY 539 GCTCCAAAGACCGCGCTCCGCCCTGAAGATGCCCGGCTTCCATTGAGAGACCAACCGCA 598
Db 604 GGACCGAGTGGGCGGCAAGCCGAGGATCGTGGGGCTTTCATGGACGGGACCAACTGCA 663
QY 599 TCGAGATCTCGAGGAGTGGAGAGGCAAGTCTTACAGACAGTACGAGGCGCGCTGG 658
Db 664 TGACGCTGGTGGACAAAGTGGCGGCGGCCAACACCTCACCATTGACTACGCTGACCAAG 723
QY 659 GCCCTACTCGACGCGCGCCCTCCAAAGTGGGCCA 695
Db 724 GCCTCTACTGACCGACCTGGACACCAACATATCGA 760

RESULT 9
BE216934 990 bp mRNA linear EST 03-JUL-2000
LOCUS EST0477 Triticum aestivum Lambda Zap Triticum aestivum cDNA clone
DEFINITION JAI SB H05_T3 5', mRNA sequence.
ACCESSION BE216934
VERSION BE216934.1 GI:9904620
KEYWORDS EST.
ORGANISM Triticum aestivum (bread wheat)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE Anderson,J.M., Williams,C.E. and Goodwin,S.B.
AUTHORS Analysis of an EST database reveals a probable CP2 resistance gene
TITLE homolog in wheat
JOURNAL Unpublished (2000)
COMMENT Contact: Anderson, J.M.
Crop Production & Pest Control Research Unit
USDA-ARS
1150 Lilly Hall, West Lafayette, IN 47907, USA
Tel: 765-494-5565
Fax: 765-496-2926
Email: janderson@purdue.edu
Seq primer: T3
High quality sequence stop: 990.
FEATURES
    source
        1..990
            /organism="Triticum aestivum"
            /mol_type="mRNA"
            /strain="P29"
            /db_xref="taxon:4565"
            /clone="JAI_SB_H05_T3"
            /tissue_type="leaf"
            /dev_stage="9 day old seedlings"
            /clone_lib="Triticum aestivum Lambda Zap"

ORIGIN
    source
        1..990
            /organism="Triticum aestivum"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZM5BMA0509120"
            /clone_lib="ZM 0.7 1.5_KB"
            /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"

Query Match 9.8%; Score 69.4; DB 10; Length 990;
Best Local Similarity 46.9%; Pred. No. 4.7; Indels 3; Gaps 1;
Matches 251; Conservative 0; Mismatches 281;

QY 32 CTTTCAGGACCAACCATCGAGGCAACCGTGAACGGCCACTACTTCAAGTCACCGGCAAGG 91
Db 211 CCGTCACCATCCAGTCATGSCAAGGCCCGCATCGGCACTTCGTGAGGCCCATGATCC 270
QY 92 GCGAGGGCAACCCCTCGAGGGCAACCGAGAGATGAAGATCGAGGTGATCGAGGGCGGCC 151
Db 271 TCGAGTCCATCGGGCTCGACTACGTGAGCAGAGAGCGAGGTCTCACGCTCGCGGACGAG 330
QY 152 CCCTGCGCCTTCGACATCTCCACATCTGTCCACCTCTCTCATGTCATCGGCTCCAAAGGCTTCA 211

```

```

Db 331 CCACCAACATCAACAGCACACTTCCGCTCCCTTCTGCTGCGGCTGCGCAACCTGG 390
QY 212 TCAAGTAGTGTGCCGCATCCCGGACTACTTCAAGACAGTCCCTCCCGAGGGCTTCACT 271
Db 391 GCGAGGCCCTCGCGCATCCGCGAGGCGCGCCATGATCGCACCAAGGCGAGGCGG 450
QY 272 GGGAGGCAACCAACCTACGAGGACGGGGTCTCTACCGCCGCCACAGACACCTCCC 331
Db 451 GCACCGGCAATGATCGAGGCGCTCCGCCACGTCGCTCCGTCATGGCGGACGTGCGTG 510
QY 332 TGGACGGGACTCGCTGTGTACAGGTGAAGATCCTGGGCAACAACTTCCCGC--CG 388
Db 511 CCTCGGAGCATGACGACGACGAGGTATTCACCTATGCCAAGAGCATCGCGCACCAT 570
QY 389 ACGGCCCCGTGATGAGAAACAGGCGCGCTGGAGCCCTTCCACCGAGATCGTGTAG 448
Db 571 ACGACTCTGTCATGACAGCAAGCAGCTCGGCGCGCTCGCGCTCCGATTCGCGCGCG 630
QY 449 AGGTGAGGCGGTGCTGCGGCGCAGTCCCTGATGAGCCCTGGAGTCCCGCGGTGCGC 508
Db 631 GCGGGGTGCGCACGCCGCGGATGCGGCTCATATGACCTCGGCTGCGACGCTGTCT 690
QY 509 ACCTGACCTGCCACCTGCACACCACTACCGCTCCAAAGACCGCCCTCCGCGCT 563
Db 691 TCGTGGCTCCGCTGCTTCAAGAGCGGGACCGCGCGCTGCGCGCGCAT 745

RESULT 10
CG332384 593 bp DNA linear GSS 26-AUG-2003
LOCUS OGVD58TV ZM 0.7 1.5 KB Zea mays genomic clone ZM5BMA0509120,
DEFINITION genomic survey sequence.
ACCESSION CG332384
VERSION CG332384.1 GI:34249650
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 593)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OGVD58TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
    Location/Qualifiers
        1..593
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZM5BMA0509120"
            /clone_lib="ZM 0.7 1.5_KB"
            /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"

ORIGIN
    source
        1..593
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZM5BMA0509120"
            /clone_lib="ZM 0.7 1.5_KB"
            /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"

Query Match 9.8%; Score 69; DB 29; Length 593;
Best Local Similarity 46.2%; Pred. No. 4.9; Indels 3; Gaps 1;
Matches 265; Conservative 0; Mismatches 305;

QY 2 GATCGCCTCCCTGCTGACCGGAGACCATGCGCCTTCAGGACCAACCATCGAGGCGCACCGTGA 61

```


Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooideae; Triticeae; Triticum.
 1 (bases 1 to 496)
REFERENCE
AUTHORS
 Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D.,
 Han, P.S., Hsia, C.C., Kang, Y., Iazo, G.R., Miller, R., Rausch, C.J.,
 Seaton, C.L. and Tong, J.C.
TITLE
 The structure and function of the expressed portion of the wheat
 genomes - Etisolated shoot cDNA library
JOURNAL
COMMENT
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105955773
 Fax: 510595818
 Email: canders@pw.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES
 source
 1..496
 Location/Qualifiers
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE0306_E02_I04"
 /tissue_type="Etisolated shoot"
 /dev_stage="Five day old seedling"
 /lab_host="E. coli SOLR"
 /clone_lib="Wheat unstressed seedling shoot cDNA library"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcorI; Site_2: XhoI; Seeds were
 surface-sterilized, germinated and grown aseptically in
 the dark at room temperature on filter paper with water,
 nystatin and ceftaxime in covered crystallization
 dishes. Shoots were harvested. The tissue, total RNA, and
 poly(A) RNA were prepared, a cDNA library was made, and
 the cDNA clones were in vivo excised to give pBluescript
 phagemids in the TJ Close lab (Choi, Close, Fenton) at the
 University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)."
ORIGIN
 Query Match 9.6%; Score 67.6; DB 13; Length 496;
 Best Local Similarity 48.3%; Pred. No. 7.5;
 Matches 221; Conservative 0; Mismatches 234; Indels 3; Gaps 1;
 QY 138 GATCGAGGGCGGCCCTCGCCCTTCCACATCTCTCCACCTCTCGCATGTACGG 197
 DB 8 GCTCGCGGACGCGCCACCATCAACAGACAACTCCGGTCCCTTGTCTGCGG 67
 QY 198 CTCGAAGGCTTATCAAGTACGTGTCCGCGATCCCGGACTTCAAGCAGTCCCTCC 257
 DB 68 CTCGCGAACCTGGGCGAGGCCCTCCGCGCATCCCGGAGGGCGCGCATGATCGTAC 127
 QY 258 CGAGGGGTTTCACTGGGAGCGCACCACTACGAGGACGGGGTCTTCTGACCGCCCA 317
 DB 128 CAGGGCGAGGCGGCACCGGCAATGTATCGAGGCGGCTCGGCACGTCCGCTCAT 187
 QY 318 CAGAGCACCTCCCTGGAAGCGGAGTCCCTGTGTGTCAGAGTGAAGATCTGGGCAACA 377
 DB 188 GGGCGACGTGCTGCTCCGAGCATGGACGATGACGAGTATTACCATATGCCAAGAG 247
 QY 378 CTTCCCGCGG---ACGGCCCGTGATGCAGACAAAGCGCGCGCTGGAGCCCTCCAC 434
 DB 248 CATTGCGGACCGTACGACTCTGTATGACAGACCAAGCAGCTCGTTCGCTCGCTGT 307
 QY 435 CGAGATGCTGTACGAGTGAAGCGGCTGTCTGCGCGCGAGTCCCTGATGCCCTGAGGTG 494
 DB 308 CCAATTGCGCGCGGCGGGTCCGACGCGCGCGATGCGGCCCTCATGATGACGTGG 367
 QY 495 CCCCAGGCTGCGGCACTGACCTGCCACCTGACACACCACTACCGCTCCCAAGAGCCCG 554

DB 368 CTGGACGGTGTCTTCGTGCGCTCGGTGTTCTTCAAGAGCGCGACCCGCGCGCGGC 427
 QY 555 CTCGCGCTGAGATGCCGCGCTTCCACTTCGAGGACC 592
 DB 428 GCGGCGCATCTGTCAGGCGGTACCCACTACAGGACC 465
RESULT 13
CA754991
LOCUS
DEFINITION
 BR030010000_PLATE_D08_60_073.ab1 OA Oryza sativa (japonica
 cultivar-group) cDNA clone BR030010000_PLATE_D08_60_073.ab1 similar
 to unknown protein [Oryza sativa (japonica cultivar-group)], mRNA
 sequence.
ACCESSION
CA754991
VERSION
 CA754991.1 GI:25799030
KEYWORDS
SOURCE
ORGANISM
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
AUTHORS
 Bohnert, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M.,
 Ferrea, H., Kawasaki, S., McCallough, A., Michalowski, C.B.,
 Palacio, C., Scarra, G., Wheeler, M. and Zepeda, G.R.
 Functional Genomics of Plant Stress Tolerance
 Unpublished (2000)
TITLE
JOURNAL
COMMENT
 Contact: Mark Fredrickson
 Department of Plant Biology
 University of Illinois
 1201 W. Gregory Dr., Urbana, IL 61801, USA
 Tel: 2172855473
 Email: bohnertlab@life.uiuc.edu.
FEATURES
 source
 1..631
 Location/Qualifiers
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="BR030010000_PLATE_D08_60_073.ab1"
 /tissue_type="roots"
 /dev_stage="3-4 weeks"
 /clone_lib="OA"
 /note="19 h 200mM NaCl"
ORIGIN
 Query Match 9.6%; Score 67.6; DB 14; Length 631;
 Best Local Similarity 47.7%; Pred. No. 7.7;
 Matches 196; Conservative 0; Mismatches 215; Indels 0; Gaps 0;
 QY 56 CCACCTACTTCAAGTGCACCGGCAAGGGGAGGAGGCAACCCCTCGAGGACCCAGGAGAT 125
 DB 8 CCCCACCCCATGTCGCGGACGCGGAGGTGGGGACCAACCGCGCGCGCGCTGGCGAC 67
 QY 126 GAAATCAGAGTGTATCGAGGGCGGCCCTTCCCTTTCGCTTCCACATCTCTGTCCACCTC 185
 DB 68 GGCACAGTTCATACCGGTCTTCGCGTCCGCGCTCGGCCACCTCGCCACGTCCA 127
 QY 186 CTGATGTAGCGGTCCAAAGGCTTTCATCAAGTACGTGTTCGGGATTCGCCGCTACTTCAA 245
 DB 128 CGGCACCGGCGACGCCCAACCTCCACCTCCATCCATCCCGCCGACGTCCA 187
 QY 246 GCAGTCCCTCCCGAGGGCTTTCACCTGGGAGCGACACCACTACGAGGACGGCGGTT 305
 DB 188 GATCGCGCGCGCGAGCTCCCGGAGCACCCCGCGCGGCGCCATGCGCGGACAGTT 247
 QY 306 CTTGACCGCCCAACGAGCACCTCCCTGGACGGGACTGCTGTGTGTACAGGTGAAGAT 365
 DB 248 CCAGGAGCGCGCAAGGACATCTCTGTGTGTCTGTGGGATCTCTTCGCGCTGCGGTG 307
 QY 366 CTTGGGCAACAACTTCCCGCGGAGCGGCCGTGTATGCAGACAAAGGCGCGCGCTGGGA 425

308 CGGCGGCTACCGCGCTCCATGTACCTCGTGTGGTCTCTCGCGCGGAGCGCG 367
 426 GCCTTCACCGAGATCGTGTAGAGGTGAGCGGCTGCTCGCGCGGAGTC 476
 368 CGCGCGGCTCCCGCTACGAGGAGCTCTACGGCGACTAGGACGAGGAGC 418

RESULT 14
 CD902570
 LOCUS G356.107G14F010919 G356 Triticum aestivum cDNA clone G356107G14,
 DEFINITION mRNA sequence.
 ACCESSION CD902570
 VERSION CD902570.1 GI:32676898
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 1 (bases 1 to 695)
 REFERENCE Genoplante.
 AUTHORS Genoplante, a major partnership french program in plant genomics
 TITLE Unpublished (2003)
 JOURNAL Contact: Genoplante
 COMMENT Genoplante
 93, rue Henri Rochefort 91025 EVRY CEDEX France
 Tel: 33 1 69 47 54 00
 Fax: 33 1 69 47 54 10
 This sequence has been generated in the framework of the french
 plant genomics programme 'Genoplante' (<http://www.genoplante.com>
 and <http://genoplante-info.inbio.gen.fr>).
 Location/Qualifiers
 1..695
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="recital"
 /db_xref="taxon:4565"
 /clone="G356107G14"
 /tissue_type="grain (356 degrees per day after
 pollination)"
 /clone_lib="G356"

ORIGIN
 Query Match 9.6%; Score 67.6; DB 14; Length 695;
 Best Local Similarity 45.9%; Pred. No. 7.8;
 Matches 269; Conservative 0; Mismatches 314; Indels 3; Gaps 1;
 45 CATCGGGGACCGTGAACGCCACTTCAAGTGCACCGGCAAGGGCGGCAACCC 104
 21 CACGGGGCAAGAGAACGTGGCGCACACCGCGGAAACCCACTTCAGATGATCCG 80
 105 CTTGAGGGGACCGAGGATGAAGATCGAGGTGATCGAGGGGGCGCCCTTCGCTTCG 164
 81 GCACGACGTGCTGAGCGGATCTCTGTGGAGTGGACGAGATCAACCACTCGCTGCC 140
 165 CTTGACATCTCTCCACCTCTCTGATGAGGTTCAGGCTTCATCAAGTACGTGTC 224
 141 GGCCTCCCGGTGACTCAAGATTCA--ACCCGTCAAGACCACTCAAGACCGTGGT 197
 225 CGGATCCCGGACTACTTCAAGACAGTCCCTCCCGAGGGGTTTCACTCGGGAGCGCAC 284
 198 GGGACGCTCAACATGCTCGGCTCGCAAGCGGTGCGGCGGCTTCCTCTCCACAG 257
 285 CACCTACGAGGAGCGGCTTCTGTACCGGCCACAGGACACTCCCTTGGACGCGACTG 344
 258 CACGAGGAGGTCTACGCGGACCCCTCCACACCCCGAGTTCGAGACCTACTGGGCAA 317
 345 CTTGTGTACAGGTGAAGATCTTGGGCAACAACTTCCCGCCGACGCGCCGCTGATGCA 404
 318 CGTCAACCCCATCGCGTGGAGTTGCTACGACGAGGCAAGAGGACGGCGGAGCGCT 377

405 GAACAAGCGCGCGCTCGGAGCCCTCCACCGAGATCGTGTACGAGGTGAGCGGTGCT 464
 378 GACCATGACTACCAACCGCGGCAACCTCGAGTGAAGTCTGTCGATCTTCAACAC 437
 465 GCGCGGACAGTCCCTGTATGGCGCTGAGTGGCGGCGGTGCGCACCTGACCTGCCACCT 524
 438 CTACGCGCGCGCATGTGATCGACGATGGCGGCTGTCAGCAACTTTCGTCGCTCAGGC 497
 525 GCACACCACTACCGCTCCAAAGAGCCCGCTCGCGCTCGAAGATGCCCGCTTCCACTT 584
 498 GTGTAGGAGAGGAGCCCTGACGGTGTACGGGACGCGCAAGCAGACGAGAGCTTCCAGTA 557
 585 CGAGGACCAACCGATCGAGATCTCGGAGAGGTGAGAGGCGCAAG 630
 558 CGTCTCCGATCTGTGTGGAGGTCTGATGAAGCTGATGAGGGGAG 603

RESULT 15
 CK159167/c
 LOCUS FGAS040564 Triticum aestivum FGAS; TaLT5 Triticum aestivum cDNA,
 DEFINITION mRNA sequence.
 ACCESSION CK159167
 VERSION CK159167.1 GI:38985053
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 1 (bases 1 to 869)
 REFERENCE Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
 Genswein,B., Grat,R., Gulick,P., Hrycan,L.D., Laroche,A.,
 Links,M.G., McCarthy,E.B., Monroy,A., Muzak,I., Nilson,D.,
 Penniket,C., Roach,J.L. and Sarhan,F.
 TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
 JOURNAL Unpublished (2003)
 COMMENT Contact: Wm L Crosby
 Bioinformatics
 University of Saskatchewan, Department of Computer Science
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,
 Saskatchewan, S7N 5A9, Canada
 Tel: 306 966 1769
 Fax: 306 966 2033
 Email: fgas.ests@usask.ca
 This sequence is the direct result of the Base calling software
 Phred (default parameters). It is the raw base calls. To aid in the
 identification of the high quality insert the software Lucy
 (default parameters) has been run on this sequence. Lucy identified
 the region (128,636).
 Plate: TaLT537 row: N column: 23.
 Location/Qualifiers
 1..869
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Wheat line PI 178383"
 /db_xref="taxon:4565"
 /lab_host="DH5 alpha"
 /clone_lib="Triticum aestivum FGAS; TaLT5"
 /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
 subtractive hybridization) cDNA library from genotype
 PI178383 cold hardened at 2 C for 21 days and 49 days
 (equal amount of cDNA pooled together before subtraction,
 tester) and subtracted against genotype Norstar cold
 hardened at 2 C for 1 day (24 H) (driver). Modified Smart
 cDNA (Clontech) priming and non-directional cloning"

ORIGIN
 Query Match 9.5%; Score 67.4; DB 14; Length 869;
 Best Local Similarity 44.7%; Pred. No. 8.5;
 Matches 263; Conservative 0; Mismatches 326; Indels 0; Gaps 0;
 18 GACCGAGACCTGCGCTTCAGGAACCACTCGAGGGCAGCGTGAACGCCACTACTTCAA 77

Wed Aug 4 09:01:06 2004

[illegible]

Search completed: August 3, 2004, 15:48:03
Job time : 3171.97 secs

Blanks Shot USPTD